

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 09:58:21 ; Search time 1807.19 Seconds

(without alignments)
-740.686 Million cell updates/sec

Title: US-09-501-714-2
Perfect score: 1376
Sequence: 1 TCTCACCAGGACTCGGACT.....TTTATTTCATATGCAAGT 1376

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenDbml:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pat:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pl4:*
11: gb_pl5:*
12: gb_pl6:*
13: gb_pl7:*
14: gb_pl8:*
15: gb_pl9:*
16: gb_pl10:*
17: gb_pl11:*
18: gb_pl12:*
19: gb_pl13:*
20: gb_pl14:*
21: gb_pl15:*
22: gb_pl16:*
23: gb_pl17:*
24: gb_pl18:*
25: gb_pl19:*
26: gb_pl20:*
27: gb_pl21:*
28: gb_pl22:*
29: gb_pl23:*
30: gb_pl24:*
31: gb_pl25:*
32: gb_pl26:*
33: gb_pl27:*
34: gb_pl28:*
35: gb_pl29:*
36: gb_pl30:*
37: gb_pl31:*
38: gb_pl32:*
39: gb_pl33:*
40: gb_pl34:*
41: gb_pl35:*
42: gb_pl36:*
43: gb_pl37:*
44: gb_pl38:*

45: gb_pl39:*
46: gb_pl40:*
47: gb_pl41:*
48: gb_pl42:*
49: gb_pl43:*
50: gb_pl44:*
51: gb_pl45:*
52: gb_pl46:*
53: gb_pl47:*
54: gb_pl48:*
55: gb_pl49:*
56: gb_pl50:*
57: gb_pl51:*
58: gb_pl52:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1328	96.5	1701	9 AB028859	AB028859 Homo sapi
2	1150.6	83.6	1159	11 HSA520137	AJ250137 Homo sapi
3	605.2	44.0	745	5 A95223	A95223 Sequence 20
4	323	23.5	89428	43 AC017242	AC017242 Drosophila
5	323	23.5	104278	35 AC005269	AC005269 Drosophila
6	263.2	19.1	145017	57 AC016926	AC016926 Homo sapi
7	171	12.4	145017	57 AC016926	AC016926 Homo sapi
8	113	8.2	332	13 HSA1842A5	Z52396 H. sapiens
9	101.4	7.4	2656	11 AF061749	AF061749 Homo sapi
10	91.8	6.7	1355	35 AF080569	AF080569 Homo sapi
11	91.4	6.6	975	40 AF063011	AF063011 Dictyostel
12	91.4	6.6	1485	9 AB014888	AB014888 Homo sapi
13	91.4	6.6	1487	9 AB015799	AB015799 Homo sapi
14	91.4	6.6	1489	40 AF075601	AF075601 Homo sapi
15	91.4	6.6	1557	40 AF060703	AF060703 Homo sapi
16	89.8	6.6	2469	9 AB015798	AB015798 Homo sapi
17	88.2	6.5	215297	32 AP001000	AP001000 Homo sapi
18	88.2	6.4	4776	1 AB018706	AB018706 Nitrosomo
19	83.4	6.1	183167	41 AC009546	AC009546 Homo sapi
20	81.2	5.9	1539	12 AB028854	AB028854 Mus muscu
21	81.2	5.9	1546	12 AF035962	AF035962 Mus muscu
22	80.2	5.8	161143	55 AC022634	AC022634 Homo sapi
23	79.6	5.8	1870	35 DMG34904	U34904 Drosophila
24	79.6	5.8	46279	42 AC014815	AC014815 Drosophila
25	79.6	5.8	112886	56 AC010661	AC010661 Drosophila
26	79.2	5.8	4874	2 AF106835	AF106835 Methylovo
27	78.8	5.7	3899	1 STD58360	U58360 Salmonella
28	78.6	5.7	6201	20 DVTIDNOT	Y07700 Drosophila
29	78.6	5.7	7412	4 FSCC5316	M99327 Torpedo cal
30	77.8	5.7	9693	12 AF092536	AF092536 Mus muscu
31	77.8	5.7	124230	12 AC005259	AC005259 Mouse BMC
32	77.6	5.6	28963	43 AC014792	AC014792 Drosophila
33	77.6	5.6	131410	42 AC008319	AC008319 Drosophila
34	77.2	5.6	1218	14 AF088983	AF088983 Mus muscu
35	76.4	5.6	1436	7 AB029547	AB029547 Schizosac
36	76.2	5.5	1140	2 AF053962	AF053962 Campyloba
37	76.2	5.5	1336	1 MAJ10152	AJ010152 Methanosa
38	75.8	5.5	282183	2 CJI1168X4	AL113077 Campyloba
39	75.4	5.5	660	8 CNS01AWK	AL113484 Botrytis
40	75.4	5.5	720	40 CNS01AWK	AL113038 Botrytis
41	75.4	5.5	1580	7 ATDNALJLK	Y11969 A. thaliana
42	75.4	5.5	12029	35 AE001405	AE001405 Plasmodia
43	74.2	5.4	4139	2 AF055368	AF055368 Vibrio ha
44	73.4	5.3	18255	42 AC013054	AC013054 Drosophila
45	73.4	5.3	82393	34 AC003923	AC003923 Drosophila

ALIGNMENTS

Db 1335 TTGTGTTAAATAAGTAAAGCATATTTATATCTGCAAGGTTTTTGTGTGCT 1394
QY 1351 TTTGTTTTTATTTTCATATGCAAGT 1376
Db 1395 TTTGTTTTTATTTTCATATGCAAGT 1420

RESULT 2
HSA250137 1159 bp mRNA PRI 07-JAN-2000
LOCUS HSA250137 1159 bp mRNA PRI 07-JAN-2000
DEFINITION Homo sapiens mRNA for ERJ3 protein (ERJ3 gene).
ACCESSION AJ250137
VERSION AJ250137.1 GI:6688202
KEYWORDS ERJ3 gene; ERJ3 protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1159)
Bies, C., Blum, R., Dudek, J. and Zimmermann, R.
Characterization of a mammalian Scj1p homolog
Unpublished
2 (bases 1 to 1159)
Zimmermann, R.
Direct Submission
Submitted (11-OCT-1999) Zimmermann R., Medizinische Biochemie und
Molekularbiologie, Universitaet des Saarlandes, Gebaeude 44,
D-66421 Homburg / Saar, GERMANY
Location/Qualifiers
1..1159
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
21..86
/gene="ERJ3"
/evidence="experimental"
21..1097
/note="ERJ3"
/function="putative role in cellular protein folding"
/note="homolog of yeast Scj1p"
/codon_start=1
/evidence="experimental"
/product="ERJ3 protein"
/protein_id="CAB65118.1"
/db_xref="GI:6688203"
/translation="MAPQNLSTFCILLILYIGAVIAGRDFYKILGYPRSAIKIDIKKA
YKRLALPHDNPDDPOAEKQDGLAAVEVLSDEKROYDTYEEGLKDGHOSSH
GLFSHFEGDFGFMGEGTRDODRNI PRGSDIVDLVLEEVYACNFYEVYKRPYA
RQAPGRKRCNCRQEMRTTOLGGRFOMTPDEYVDCDECPNKLVAERTLEVEIPEYRD
GMEYPTIGGEPRVDEPGDLRPRIKVVRHPIFERGGDDLVTNVTLSLVESLVGFEMD
LTHLDHKVHISDKITTPGAKIMKKEGLPNPDNNINIKSLIITPDVDFPKQLTEE
AREGIKOLLKQGSVKVYNGLOGY"
21..1097
/gene="ERJ3"
87..1094
/gene="ERJ3"
/note="homolog of yeast Scj1p"
/evidence="experimental"
/product="ERJ3"
BASE COUNT 355 a 216 c 314 g 274 t
ORIGIN

Query Match 83.6%; Score 1150.6; DB 11; Length 1159;
Best local similarity 99.5%; Pred. No. 1.8e-258;
Matches 1153; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 176 GACCCGGGACAGAGAACCATGGCTCCGACAACTGACACCTTTGGCTGTGCTGC 235
Db 1 GAACCCGGGACAGAGAACCATGGCTCCGACAACTGACACCTTTGGCTGTGCTGC 60
QY 236 TATACCTCATCGGGGGGCGGTATTCGCGAGAGATTCTTAAGATTCTGGGGGTGCTC 295
|||||

Db 61 TATACCTCATCGGGGGGCGGTATTCGCGAGAGATTCTTAAGATTCTGGGGGTGCTC 120
QY 296 GAAAGCTCTCTATTAAGAGATTTAAAAAGCCTATAGAAAAGCTGACCTGATC 355
Db 121 GAAAGCTCTCTATTAAGAGATTTAAAAAGCCTATAGAAAAGCTGACCTGATC 180
QY 356 CCGACCGGAACCCGATGATCCACAAAGCCAGAGAAATCCGAGATCTGGGTCTCTT 415
Db 181 CCGACCGGAACCCGATGATCCACAAAGCCAGAGAAATCCGAGATCTGGGTCTCTT 240
QY 416 ATGAGTTCTGTCCAGATAGTGAAGAAAGAAACAGTACGATCTTATGTGAAGAGAT 475
Db 241 ATGAGTTCTGTCCAGATAGTGAAGAAAGAAACAGTACGATCTTATGTGAAGAGAT 300
QY 476 TAAAGATGGTCATCAGAGCTCCCATGAGACATTTTTCACACTCTTGGGATTTTG 535
Db 301 TAAAGATGGTCATCAGAGCTCCCATGAGACATTTTTCACACTCTTGGGATTTTG 360
QY 536 GTTTCATGTTTGGAGAAACCCCTGTCAGCAAGACAAATATTCAGAGAGAGATGATA 595
Db 361 GTTTCATGTTTGGAGAAACCCCTGTCAGCAAGACAAATATTCAGAGAGAGATGATA 420
QY 596 TTATTGTAGATCTAGAGTCACTTTGGAAGAGATATGACAGAAATTTTGTGGAAGTAG 655
Db 421 TTATTGTAGATCTAGAGTCACTTTGGAAGAGATATGACAGAAATTTTGTGGAAGTAG 480
QY 656 TTAGAAACAAACCTGTGGCAAGGCGAGCTCCGTGGCAAAAGGAAATGTCGGAAG 715
Db 481 TTAGAAACAAACCTGTGGCAAGGCGAGCTCCGTGGCAAAAGGAAATGTCGGAAG 540
QY 716 AGATGCGGACCAACCACTGAGGCGCTGCGGCTTCCAAATGACCAAGAGAGTGTGCGG 775
Db 541 AGATGCGGACCAACCACTGAGGCGCTGCGGCTTCCAAATGACCAAGAGAGTGTGCGG 600
QY 776 ACGAATGCCCTTAATGTCAAACTAGTGAATGAAGAACACGCTGGAAGTAGAAATAGAGC 835
Db 601 ACGAATGCCCTTAATGTCAAACTAGTGAATGAAGAACACGCTGGAAGTAGAAATAGAGC 660
QY 836 CTGGGGTGAGAGACGCGATGAGTACCCCTTATTTGGAAGAGTGTGAGCTGAGTGTG 895
Db 661 CTGGGGTGAGAGACGCGATGAGTACCCCTTATTTGGAAGAGTGTGAGCTGAGTGTG 720
QY 896 GGGAGCCTGGAGATTTACGCTTCCGAATCAAAGTGTGCAAGCACCACATATTTAAAGGA 955
Db 721 GGGAGCCTGGAGATTTACGCTTCCGAATCAAAGTGTGCAAGCACCACATATTTAAAGGA 780
QY 956 GAGAGATGATTTTACAAATGTCACAGTCAGTCATTAGTTGAGTCACTGGTGGCTTTG 1015
Db 781 GAGAGATGATTTTACAAATGTCACAGTCAGTCATTAGTTGAGTCACTGGTGGCTTTG 840
QY 1016 AGATGATATTTACTACCTGGATGTCACAAAGTACATATTTCCCGGATTAAGTACCA 1075
Db 841 AGATGATATTTACTACCTGGATGTCACAAAGTACATATTTCCCGGATTAAGTACCA 900
QY 1076 GGGCAGAGAGCAANTATGGAAGAAGGGAGGCTCCCACTTTGACAAACATA 1135
Db 901 GGGCAGAGAGCAANTATGGAAGAAGGGAGGCTCCCACTTTGACAAACATA 960
QY 1136 TCAAGGCTCTTTGATATACCTTTGATGATGATTTTCCAAAAGACAGTATACAGAG 1195
Db 961 TCAAGGCTCTTTGATATACCTTTGATGATGATTTTCCAAAAGACAGTATACAGAG 1020
QY 1196 AAGCAGAGAGATATCAACAGCTACTGAAACAGGGCTAGTCAGAGGTATATACATG 1255
Db 1021 AAGCAGAGAGATATCAACAGCTACTGAAACAGGGCTAGTCAGAGGTATATACATG 1080
QY 1256 GACTGCAAGATATTTGAGATGATTAATTTGACTTTGTTAAATTAAGTAAGCG 1315
Db 1081 GACTGCAAGATATTTGAGATGATTAATTTGACTTTGTTAAATTAAGTAAGCG 1140
QY 1316 ATATTATTTATCTGCAAGG 1334
Db 1141 ATATTATTTATCTGCAAGG 1159

RESULT 3
LOCUS A95223 745 bp DNA PAT 26-JAN-2000
DEFINITION Sequence 20 from Patent WO928461.
ACCESSION A95223
VERSION A95223.1 GI:6779297
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 745)
AUTHORS Danen-Van, O.A. and Noteborn, M.H.
TITLE METHODS AND MEANS FOR INDUCING APOPTOSIS BY INTERFERING WITH
BIP-LIKE PROTEINS
JOURNAL Patent: WO 928461-A, 10-JUN-1999;
LEAD B V (NL); DANEN VAN OORSCHOT ASTRID ADRI (NL)
FEATURES
Source 1..745
Location/Qualifiers
BASE COUNT 237 a 128 c 184 g 177 t 19 others
ORIGIN
Query Match 44.0%; Score 605.2; DB 5; Length 745;
Best Local Similarity 89.6%; Pred. No. 3, 6e-131;
Matches 660; Conservative 0; Mismatches 74; Indels 3; Gaps 2;
QY 590 GTGATATTATTTGATCTGAAAGTCACCTTGGAGAATATATGAGAAATTTGTGG 649
DB 1 GTATATATTATGATCTGAAAGTCACCTTGGAGAATATATGAGAAATTTGTGG 60
QY 650 AAGTAGTTAGAAACCAACCTGTGGCAAGGAGCTCTGTCGAACGGAAGTGAATTTGTC 709
DB 61 AAGTAGTTAGAAACCAACCTGTGGCAAGGAGCTCTGTCGAACGGAAGTGAATTTGTC 120
QY 710 GGCAGAGATGCGGACCAACCTGTGGCCCTGTGCAATGACCCAGAGGTGG 769
DB 121 GGCAGAGATGCGGACCAACCTGTGGCCCTGTGCAATGACCCAGAGGTGG 180
QY 770 TCTGCCACCAATCCCTTAATGTCAAACTAGTGAATGAGACGACCTGGAAATGAAA 829
DB 181 TCTGCCACCAATCCCTTAATGTCAAACTAGTGAATGAGACGACCTGGAAATGAAA 240
QY 830 TAGAGCTGGGTGAGAGAGCGCATGAGTACCCCTTTATTTGAGAAGGTGACCTCAG 889
DB 241 TAGAGCTGGGTGAGAGAGCGCATGAGTACCCCTTTATTTGAGAAGGTGACCTCAG 300
QY 890 TGGATGGGAGCCTGGAGATTTACGGTTCCGAATCAAGTTGTCAAGCAACCAATATTG 949
DB 301 TGGATGGGAGCCTGGAGATTTACGGTTCCGAATCAAGTTGTCAAGCAACCAATATTG 360
QY 950 AAGAGGAGAGATGATTTGTACAAATGTGACATCTCATTTAGTTGAGTCACTGGTG 1009
DB 361 AAGAGGAGAGATGATTTGTACAAATGTGACATCTCATTTAGTTGAGTCACTGGTG 420
QY 1010 GCTTTGAGATGATTTACTCACTTGTGATGTCACAAAGTATATTTCCCGGAT -AAG 1068
DB 421 GCTTTGAGATGATTTACTCACTTGTGATGTCACAAAGTATATTTCCCGGAT -AAG 480
QY 1069 ATCCAGAGCCAGAGACCGAATATGGAAGAAAGGGGCTCCCAACTTTGCAAC 1128
DB 481 ATCCAGAGCCAGAGACCGAATATGGAAGAAAGGGGCTCCCAACTTTGCAAC 540
QY 1129 AACCAATATCAAGGCTCTTTGATTAATCACTTTTGTGATTTTCCAAAGAACAGTTA 1188
DB 541 AACCAATATCAAGGCTCTTTGATTAATCACTTTTGTGATTTTCCAAAGAACAGTT 600
QY 1189 ACAGAGAGAGCGAGAGAGTATCAACAGTACTGAAACAGGGTCACTGAGAGAGTA 1248
DB 601 ACAGAGAGAGCGAGAGAGTATCAACAGTACTGAAACAGGGTCACTGAGAGAGTA 658

QY 1249 TACAAATGACTCGAAGATTTTGAGAGTGAATTAATTTGACTTTGTTAAATTAAGTGA 1308
DB 659 TNCAAATGACCGAAGCAATTTGAAANTGAATTAATTTGTTAAATTAATTAATTAATTA 718
QY 1309 ATAGAGATATTTATTTA 1325
DB 719 NCCATATTTATNTATCA 735
RESULT 4
LOCUS AC017242/c 89428 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017242
VERSION AC017242.1 GI:6553744
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 89428)
AUTHORS Adams, M. and Venter, J.C.
TITLE Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
JOURNAL Direct Submission
COMMENT This sequence was identified as CDM:1020998 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source 1..89428
Location/Qualifiers
BASE COUNT 24712 a 19587 c 20235 g 24894 t
ORIGIN
Query Match 23.5%; Score 323; DB 43; Length 89428;
Best Local Similarity 59.4%; Pred. No. 3, 7e-65;
Matches 586; Conservative 0; Mismatches 392; Indels 9; Gaps 2;
QY 226 CTGTGCTCTTACCTCACTGCGGGGCTGATTCGCGGACGATTTCTATTAAGATCTTG 285
DB 1420 CAGCTGTCTCTTCTGCTGTGAGAGATCATTTTCCCGGCGGACTTCTACAAATCTG 1361
QY 286 GGGTGCCTCGAAGTCCCTTAATTAAGATTAATTAAGAGCCTATAGAACTAGCCCTG 345
DB 1360 AACGTAAAGAAACCGCAACGCAAGAGTGAAGAAAGCCTATAGAGCCTTGGCCAG 1301
QY 346 CAGCTTCATCCGACCGGAACCTGATGATCCACAAGCCAGAGAAATTTCCAGATCTG 405
DB 1300 GAGCTGATCCGATTAAGAAACAGACGACCGGATGCTCCACAATTTCCAGACTG 1241
QY 406 GGTGCTGTATAGAGTCTGTCAATAGTGAAGAAAGGAAACAGTACATTTTANGT 465
DB 1240 GGAGGGCTTCAAGAGTCTCTTCAATCCGGAACGAAGAGCTTACGACCGTGGCGC 1181
QY 466 GAAGAAGATT---AAAAGATGTCATCAGAGCTCCATGGAGACATTTTTCACACTG 522
DB 1180 GAGGAATGCTTCAAGAAAGAGGAGCATGATGATCACGGTGTGATCCGTTCTTACGCTC 1121
QY 523 TTTGGGATTTTGTGCTTTCATGTTTGGAGAAACCCCTGTCAGCAAGACGAATATTCCA 582
DB 1120 TTTGGGACTTCGCGCTTTCATCTGCTGCTGATGGCCAGCAGCAAG-----ATGCTCG 1067
QY 583 AGAGAAATGATTTATTTATGATCTAGAACTCACTTTGGAAGAAAGTATATGACGAAT 642
DB 1066 CGAGCGCCGATATCGTATGAGCTTTCCTCCGAGAGAGCTATACTCCGGAAC 1007


```

OY 643 TTGTGAGAGTACTTAGTAAACAAACCTGTGGCAAGCGCTCCTGGCAACGGAAGTGC 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1006 TTGTGGAATTTGTAGGAACAAGCTGTACAAACCCCTCAGGACACGAGAAATGCG 947
OY 703 AATTGTGCGCAGAGATGCGGACCAACCCAGCTGGCCCTTGCGGCTTCCAAATGACCCAG 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 946 AACTGCCGCGAGAGATGTGTCACCCGGAACCTTGAGACCCGGGGCTTCCAGATGATCCAA 887
OY 763 GAGGTGCTGTCGCGAAGATGCCCTAATGTCAAACTAGTGAAAGAAAGCAAGCTGGAA 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 886 CAAACGCTGTGCGACGAGTGTCCCAAGGTAGTCAAGCAAGAGGAGCCACATTTGAG 827
OY 823 GTTGAATAGACCTGGGGTGAAGAGCGCATGAGTACCCCTTATGTGAAGAGTGAG 882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 826 ATCGAGGTGGAACAGGCAATGTCGACGGCCAGAGCGAGTGTGCGGCGAGGGGAG 767
OY 883 CCTCAGGTGATGGAGGACCTTGAGATTTACGGTCCGAATCAAAAGTGTCAAGCAACCA 942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 766 CCACATATGATGGAGGACCCGCGACCTCATTTGTGCGGTTCAACAATGCCGATCCG 707
OY 943 AATATTGAAGAGAGAGATGATTTGTACCAAAATGTGACAGTCTCATTTAGTTGATCA 1002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 706 CGATTCCTGCGCAAGATGATGTATGTACAGAGAGTGCATGTCGAGGATGCC 647
OY 1003 CTGTGTGGCTTTAGATGATATTACTCACTTGGATGTGTACAGAGTACATATTTCCCG 1062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 646 CTGTGTGGCTTGTCCATGTGATCAAGCACTTGTATGATGACACCTGTGCGGCTCACGAGA 587
OY 1063 GATTAATCATCCAGCGCAGAGGAGGAANTANTGGAAGAAAGGGAGGCGCCCAACTTT 1122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 586 GAGAGAGTAACTGTGGCCCGGCGCTAGATATCCGCAAGAGGGCAGGGCATGCCAATCTT 527
OY 1123 GACACACAAATATCAAGGGCTTTGTATATCACTTTTGTGATTTTCCAAAAGAA 1182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 526 GAGAACAAACAACCTCACCGCACTGTATCATCACTTGATGTGGATGCCCAAGAG 467
OY 1183 CAGTTACAGAGAGAGCGAGAGAGGT 1209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 466 GATCTTACGAGAGAGAGAGAGAGGT 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AC005269 104278 bp DNA INV 07-JUL-1998
LOCUS Drosophila melanogaster DNA sequence (Pls DS00764 (D273) and
DEFINITION DS00501 (D274)), complete sequence.
AC005269 AC004571 AC004572
VERSION AC005269.1 GI:3293205
KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in tet from p1 clones DS00764
ORGANISM (D273) and DS00501 (D274)) DNA.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 104278)
Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Abpayani,A., Arcinas,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Humastli,S.R., Katra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zietan,L.L. and Kimmel,B.
Sequence of Drosophila chromosome, region 21C3-21C7
Unpublished (1997)
2 (bases 1 to 104278)
Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Abpayani,A., Arcinas,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Humastli,S.R., Katra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zietan,L.L. and Kimmel,B.
Sequence of Drosophila chromosome, region 21C3-21C7
Unpublished (1997)
2 (bases 1 to 104278)

```

```

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zietan,L.L. and Kimmel,B.E.
Direct Submission
Submitted (07-JUL-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://fruitfly.berkeley.edu/sequence/) or send
email to drosophila@mcg.lbl.gov.
Library locations: 135-8, 21-6.
Location/Qualifiers
1. 104278
FEATURES
source
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/map="21C3-21C7"
/clone="Pls DS00764 (D273) and DS00501 (D274)"
/note="DS00764 (d273) extends from bp 1 to bp 38,674 and
DS00501 (d274) extends from bp 20,505 to bp 104,278."
BASE COUNT 28186 a 23704 c 23278 g 29110 t
ORIGIN
Query Match 23.5%; Score 323; DB 35; Length 104278;
Best Local Similarity 59.4%; Pred. No. 3.7e-65;
Matches 586; Conservative 0; Mismatches 392; Indels 9; Gaps 2;
OY 226 CTGTGTGCTATACATCATCGGGCGGTAGTCCGAGAGATTTCTAATGATCTTG 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16821 CAGCTGTCCCTCTGCTGTGGAGAGATCATTTCCGGCGGAGCTCTCAAAATATCTG 16880
OY 286 GGGGTGCTCGAGAGTGCCTCTATTAAGGATTTAAAAAGCCTTATAGAAACTAGCCCTG 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16881 AACTTAAGAAGAAAACGCAACACGAAAGTGAAGTGAATGCCCTTTGGCCAAAG 16940
OY 346 CAGCTTATCCCGACCGGAACCTGATATCCAAACCCAGAGAAATTCAGAGATCTG 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16941 GAGCTGCATCCGATTAAGAACAGAGACCCGATGCTTCACAAAGTTCCAGAGACTG 17000
OY 406 GTGCTGCTTATAGAGTTCGTGTAGATGAGAAAGCAAGTACGATCTATATGCT 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17001 GGAGCGGCTACGAAGTCTCTCCAAATCCGAGCAAAAGGAAGACTACGACCGTGGCG 17060
OY 466 GAAGAGAGATT--AAAAGATGTCATCAGAGCTCCCATGAGACATTTTTCACACTTC 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17061 GAGGAATGCTTCAAGAGAGAGGAGGATGATGATCAGGTGTGCTGCTTACGCTTC 17120
OY 523 TTGTGGGATTTGTGTTTCATGTTTGGAGGAACCCCTGTAGCAAGACAGAAATATTCGA 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17121 TTGTGGGACTTCGCTTCACTTCGTGTGTAGTGGCCAGAGCAAG-----ATGTCGCG 17174
OY 583 AGAGAGATATATTTATTTAGATCTAGATCACTTGTGAAGAGATATATGCAAGAT 642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17175 CGAGCGCCCATATCGTAATGAGACTTGTACGTTCCCTGAGAGAGCTATATCTCGGAAC 17234
OY 643 TTGTGGAAGTATTAGAAACAAACCTGTGGCAAGGAGGCTCTGCAAAAGCAAGTGC 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17235 TTGTGGAATTTGTGAGGAAAGCTGTAAAGAAACCCCTTCACGACACCAAGAAATGCG 17294
OY 703 AATTGTGCGCAAGAGATGCGGACCAACCTGAGGCGCTCTGCAAAAGCAAGTGC 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17295 AACTGCCGCGAGAGAGTGTACCCGGAACCTTGAGACCCGGGGCTTCCAGATGATCCAA 17354
OY 763 GAGGTGCTGCGAGCAATGCCCTAATGTCAAACTAGTGAAGAAAGCAAGCTGGAA 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17355 CAGAGGCTGTGCGACGAGTGTCCCAACGTAAGCTAGTCAACGAGAGGCGCACATTTGGAG 17414

```

OY 823 GTAGAAATAGACCTGGGGTAGAGACGCGCATGTGATACCCCTTTATTTGAGAAAGTGC 882
 Db 17415 ATCAGAGGTGAACAGAGGCATGTCGACGCCAAGACGAGGTTCGTGGCCGAGGGAG 17474
 OY 883 CCTACGCGATGGGGGAGCCTTGAGATTTCAGGTTCCGAATCAAGTTGTCACGCCCA 942
 Db 17475 CCACATATCGATGGGGAGCCCGCGACCTCTATTGTGGGGTTCAACAAATGCCCATCCG 17534
 OY 943 ATATTGAAAGAGAGAGATGATTTGTACACAAATGTGACATCTCATTTAGTTAGTCA 1002
 Db 17535 CGATTCCGCGCAGAAATGATGATCTGTACACGAACGTGACATCATCTGCAGAGATCC 17594
 OY 1003 CTGTTGGCTTTGAGATGATATTACTGCTGATGATGTCACAAAGTATTTCCCGG 1062
 Db 17595 CTTGTCGGCTTCTCCATGGATATCAGCCTTGTGATGACACCTGTGCCGTCACGAGA 17654
 OY 1063 GATAGATATCACAGGAGGAGGAGCAANTANTGAAAGGAGGAGGCTCCCACTTT 1122
 Db 17655 GAGAGGTATACGTGGCCCGGCTAGAAATCCGCAAGAGGCGGCGCATGCCACTTT 17714
 OY 1123 GACACACACATATATCAAGGCTCTTTGATATCACTTTGATGATGATTTCCAAAGAA 1182
 Db 17715 GAGAACACAAACCTCACCGGACCTGTACATCACTTCATGTGAGATGCCCAAGAG 17774
 OY 1183 CAGTTAACAGAGGAGGAGGAGGAGT 1209
 Db 17775 GATCTTACGAGGAGGAGGAGGAGGAGT 17801

RESULT 6
 AC016926
 LOCUS Homo sapiens clone RP11-35G16, *** SEQUENCING IN PROGRESS ***, 60
 DEFINITION unordered pieces.
 ACCESSION AC016926 GI:7007843
 VERSION AC016926.3
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 145017)

Muzny,D.M., Adams,C., Bailey,M., Barbite,J., Blankenburg,K.,
 Boretta,B., Bouck,J., Bowler,S., Brooks,A., Bunay,C., Bunac,C.,
 Burdett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
 David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
 Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
 Forcum-Tansey,J., Frenzt,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
 Guenava,M., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
 Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,X., Jones,M.,
 Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
 Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
 Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
 Morris,S., Nash,S., Nelson,A., McLeod,M.P., Mei,G., Morgan,M.,
 Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
 Oulles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
 Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugan,R.,
 Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabnah,M.,
 Wallington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
 Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
 Gibbs,R.

TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 145017)
 AUTHORS Worley,K.C.

TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Feb 19, 2000 this sequence version replaced gi:6671836.

COMMENT Center: Baylor College of Medicine
 Genome Center
 Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc_help@bcm.tmc.edu
 Project Information
 Center Project name: HMR
 Center clone name: RP11-35G16
 Summary Statistics
 Sequencing vector: M13; 108821
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.980611
 Consensus quality: 90208 bases at least Q40
 Consensus quality: 108595 bases at least Q30
 Consensus quality: 118311 bases at least Q20
 Estimated insert size: 145017; agarose-IP estimation
 Estimated insert size: 134603; sum-of-contigs estimation
 Quality coverage: 1.4x in Q20 bases; agarose-IP estimation
 Quality coverage: 1.5x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 60 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	784	783: contig of 783 bp in length
*	804	gap of unknown length
*	1652	contig of 843 bp in length
*	1653	gap of unknown length
*	1672	contig of 844 bp in length
*	1673	gap of unknown length
*	2517	contig of 847 bp in length
*	2537	gap of unknown length
*	3384	contig of 1038 bp in length
*	3404	gap of unknown length
*	4442	contig of 1038 bp in length
*	4461	gap of unknown length
*	5249	contig of 788 bp in length
*	5250	gap of unknown length
*	5270	contig of 867 bp in length
*	6137	gap of unknown length
*	6157	contig of 1497 bp in length
*	7654	gap of unknown length
*	7674	contig of 817 bp in length
*	8491	gap of unknown length
*	8511	contig of 1918 bp in length
*	10429	gap of unknown length
*	10449	contig of 794 bp in length
*	11243	gap of unknown length
*	11263	contig of 1300 bp in length
*	12563	gap of unknown length
*	12582	contig of 778 bp in length
*	12583	gap of unknown length
*	13361	contig of 935 bp in length
*	13381	gap of unknown length
*	14316	contig of 1415 bp in length
*	14336	gap of unknown length
*	14336	contig of 1415 bp in length
*	15751	gap of unknown length
*	15771	contig of 1438 bp in length
*	17209	gap of unknown length
*	17229	contig of 1056 bp in length
*	18285	gap of unknown length
*	18305	contig of 883 bp in length
*	19188	gap of unknown length
*	19208	contig of 1929 bp in length
*	21137	gap of unknown length
*	21157	contig of 1533 bp in length
*	22690	gap of unknown length
*	22710	contig of 3037 bp in length
*	25747	gap of unknown length
*	25766	contig of 1481 bp in length
*	27267	gap of unknown length
*	27268	contig of 1365 bp in length
*	28633	gap of unknown length
*	28653	contig of 1780 bp in length
*	30432	gap of unknown length
*	30433	gap of unknown length

30453 32234: contig of 1782 bp in length
32235 32254: gap of unknown length
32255 34215: contig of 1961 bp in length
34216 34235: gap of unknown length
34236 35966: contig of 1731 bp in length
35967 35986: gap of unknown length
37169 37189: contig of 1183 bp in length
37170 37189: gap of unknown length
37190 39327: contig of 2138 bp in length
39328 39347: gap of unknown length
39348 40797: contig of 1450 bp in length
40798 40817: gap of unknown length
40818 42978: contig of 2161 bp in length
42979 42998: gap of unknown length
42999 45518: contig of 2520 bp in length
45519 45538: gap of unknown length
45539 47419: contig of 1881 bp in length
47420 47439: gap of unknown length
47440 49427: contig of 1988 bp in length
49428 49447: gap of unknown length
49448 51749: contig of 2302 bp in length
51750 51769: gap of unknown length
51770 53862: contig of 2093 bp in length
53863 53882: gap of unknown length
53883 56866: contig of 2984 bp in length
56867 56886: gap of unknown length
56887 59350: contig of 2464 bp in length
59351 59370: gap of unknown length
59371 61846: contig of 2476 bp in length
61847 61866: gap of unknown length
61867 63930: contig of 2064 bp in length
63931 63950: gap of unknown length
63951 65797: contig of 1847 bp in length
65798 65817: gap of unknown length
65818 69175: contig of 3358 bp in length
69176 69195: gap of unknown length
69196 72751: contig of 3556 bp in length
72752 72771: gap of unknown length
72772 75841: contig of 3070 bp in length
75842 75861: gap of unknown length
75862 78423: contig of 2562 bp in length
78424 78443: gap of unknown length
81402 81422: contig of 2959 bp in length
81403 81423: gap of unknown length
81424 84125: contig of 2703 bp in length
84126 84145: gap of unknown length
84146 87688: contig of 3543 bp in length
87689 87708: gap of unknown length
87709 91413: contig of 3705 bp in length
91414 91433: gap of unknown length
91434 95783: contig of 4350 bp in length
95784 95803: gap of unknown length
95804 98920: contig of 3117 bp in length
98921 98940: gap of unknown length
98941 102954: contig of 4014 bp in length
102955 102974: gap of unknown length
102975 107971: contig of 4997 bp in length
107972 107991: gap of unknown length
107992 111651: contig of 3660 bp in length
111652 111671: gap of unknown length
111672 115722: contig of 4051 bp in length
115723 115742: gap of unknown length
115743 121800: contig of 6058 bp in length
121801 121820: gap of unknown length
121821 127723: contig of 5903 bp in length
127724 127743: gap of unknown length
127744 133403: contig of 5660 bp in length
133404 133423: gap of unknown length
133424 139012: contig of 5589 bp in length
139013 145017: contig of 5985 bp in length.
139033 145017: contig of 5985 bp in length.
Location/Qualifiers
1. 145017

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-35G16"
BASE COUNT 40495 a 30419 c 31541 g 41252 t 1310 others
ORIGIN
Query Match 19.1%; Score 263.2; DB 57; Length 145017;
Best Local Similarity 98.9%; Pred. No. 3.5e-51;
Matches 265; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TCTCACCAGGAGTGGGAGCTCCCGGAAGTGAGCCGAGAGAGGGGGCTAGTAGCTG 60
Db 128202 TCTCACCAGGAGTGGGAGCTCCCGGAAGTGAGCCGAGAGAGGGGGCTAGTAGCTG 128261
QY 61 TCTCTGGCGAGGAGGAGACCCCGCCCGCGGTGTGAGGGCGCCCTCAGAGGCCGG 120
Db 128262 TCTCTGGCGAGGAGGAGACCCCGCCCGCGGTGTGAGGGCGCCCTCAGAGGCCGG 128321
QY 121 GTGGGCTGGCAGCGCCGCGCGCGCGCGAGAGGCTGTGAGAGCTGTGTGACACAGACC 180
Db 128322 GTGGGCTGGCAGCGCGCGCGCGCGAGAGGCTGTGAGAGCTGTGTGACACAGACC 128381
QY 181 CGGAGACAGAGAACCATGCTCCGACAGACCTTTTCTGCTGCTATAC 240
Db 128382 CGGAGACAGAGAACCATGCTCCGACAGACCTTTTCTGCTGCTATAC 128441
QY 241 CTCATCGGGGGGGTGTGATTCGCGAGCAG 268
Db 128442 CTCATCGGGGGGGTGTGATTCGCGAGCAG 128469
RESULT 7
AC016926/c DNA 19-FEB-2000
LOCUS Homo sapiens clone RP11-35G16, *** SEQUENCING IN PROGRESS ***, 60
DEFINITION Unordered pieces.
ACCESSION AC016926 GI:7007843
VERSION AC016926.3 GI:7007843
KEYWORDS HTG; HTGS; PHASEL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145017)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunz,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantiz,P., Ganesb,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,M., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondolewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichter,A., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,K., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugrany,R.,
Tabbor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabish,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145017)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Feb 19, 2000 this sequence version replaced gi:6671836.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMSR
Center clone name: RP11-35G16
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 90208 bases at least Q40
Consensus quality: 108595 bases at least Q30
Consensus quality: 118331 bases at least Q20
Estimated insert size: 145017; agarose-fp estimation
Estimated insert size: 134603; sum-of-ctgigs estimation
Quality coverage: 1.4x in Q20 bases; agarose-fp estimation
Quality coverage: 1.5x in Q20 bases; sum-of-ctgigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
784 783: contig of 783 bp in length
804 803: gap of unknown length
1653 1652: contig of 849 bp in length
1673 1672: gap of unknown length
2517 2516: contig of 844 bp in length
2537 2536: gap of unknown length
3383 3383: contig of 847 bp in length
3404 3403: gap of unknown length
4441 4441: contig of 1038 bp in length
4461 4461: gap of unknown length
5249 5249: contig of 788 bp in length
5269 5269: gap of unknown length
6136 6136: contig of 867 bp in length
6156 6156: gap of unknown length
7653 7653: contig of 1497 bp in length
7673 7673: gap of unknown length
8490 8490: contig of 817 bp in length
8510 8510: gap of unknown length
10428 10428: contig of 1918 bp in length
10449 10448: gap of unknown length
11242 11242: contig of 794 bp in length
11263 11262: gap of unknown length
12562 12562: contig of 1300 bp in length
12583 12582: gap of unknown length
13360 13360: contig of 778 bp in length
13380 13380: gap of unknown length
13361 13361: contig of 935 bp in length
14316 14315: gap of unknown length
14335 14335: gap of unknown length
15750 15750: contig of 1415 bp in length
15770 15770: gap of unknown length
17208 17208: contig of 1438 bp in length
17228 17228: gap of unknown length
18284 18284: contig of 1056 bp in length
18304 18304: gap of unknown length
19187 19187: contig of 883 bp in length
19207 19207: gap of unknown length
21136 21136: contig of 1929 bp in length
21156 21156: gap of unknown length
22689 22689: contig of 1533 bp in length
22709 22709: gap of unknown length
25746 25746: contig of 3037 bp in length
25766 25766: gap of unknown length
27247 27247: contig of 1481 bp in length
27267 27267: gap of unknown length
27268 27268: contig of 1365 bp in length

28633 28652: gap of unknown length
28653 30432: contig of 1780 bp in length
30433 30452: gap of unknown length
30453 32234: contig of 1782 bp in length
32235 32254: gap of unknown length
34215 34215: contig of 1961 bp in length
34216 34235: gap of unknown length
34236 35966: contig of 1731 bp in length
35967 35986: gap of unknown length
35987 37189: contig of 1183 bp in length
37190 39327: contig of 2138 bp in length
39328 39347: gap of unknown length
39348 40797: contig of 1450 bp in length
40798 40817: gap of unknown length
40818 42978: contig of 2161 bp in length
42979 42998: gap of unknown length
42999 45518: contig of 2520 bp in length
45519 45538: gap of unknown length
45539 47419: contig of 1881 bp in length
47420 47439: gap of unknown length
47440 49427: contig of 1988 bp in length
49428 49447: gap of unknown length
49448 51749: contig of 2302 bp in length
51750 51769: gap of unknown length
51770 53862: contig of 2093 bp in length
53863 53882: gap of unknown length
53883 56866: contig of 2984 bp in length
56867 56886: gap of unknown length
56887 59350: contig of 2464 bp in length
59351 59370: gap of unknown length
59371 61846: contig of 2476 bp in length
61847 61866: gap of unknown length
61867 63930: contig of 2064 bp in length
63931 63950: gap of unknown length
63951 65797: contig of 1847 bp in length
65798 65817: gap of unknown length
65818 69175: contig of 3358 bp in length
69176 69195: gap of unknown length
69196 72751: contig of 3556 bp in length
72752 72771: gap of unknown length
72772 75841: contig of 3070 bp in length
75842 75861: gap of unknown length
78423 78443: contig of 2562 bp in length
78444 78444: gap of unknown length
81402 81402: contig of 2959 bp in length
81403 81422: gap of unknown length
81423 84125: contig of 2703 bp in length
84126 84145: gap of unknown length
84146 87688: contig of 3543 bp in length
87689 87708: gap of unknown length
87709 91413: contig of 3705 bp in length
91414 91433: gap of unknown length
91434 95783: contig of 4350 bp in length
95784 95803: gap of unknown length
95804 98920: contig of 3117 bp in length
98921 98940: gap of unknown length
98941 102954: contig of 4014 bp in length
102955 102974: gap of unknown length
102975 107971: contig of 4997 bp in length
107972 107991: gap of unknown length
107992 111651: contig of 3660 bp in length
111652 111671: gap of unknown length
111672 115722: contig of 4051 bp in length
115723 115742: gap of unknown length
115743 121800: contig of 6058 bp in length
121801 121820: gap of unknown length
121821 127723: contig of 5903 bp in length
127724 127743: gap of unknown length
127744 133403: contig of 5660 bp in length
133404 133423: gap of unknown length
133424 138012: contig of 5589 bp in length
139013 139032: gap of unknown length

FEATURES	* 139033	145017: contig of 5985 bp in length.
SOURCE	Location/Qualifiers	
	1. 145017	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="Rp11-35G16"	
BASE COUNT	40495 a	30419 c 31541 g 41252 t 1310 others
ORIGIN		
Query Match	12.4%;	Score 171; DB 57; Length 145017;
Best Local Similarity	100.0%;	Pred. No. 1.2e-29;
Matches 171;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY 1206	AGGATTCAAACAGCTACTGAAACAAAGGCTCACTGCAGAGGATATCAATGAGCTGCACAG	1265
Db 66241	AGGTTCACAAACAGCTACTGAAACAAAGGCTCACTGCAGAGGATATCAATGAGCTGCACAG	66182
QY 1266	ATATTGACAGCGCAATAAATTTGCACTTTGTTTAAATAAGTCATAATGAGATATTATTA	1325
Db 66181	ATATTGACAGCGCAATAAATTTGCACTTTGTTTAAATAAGTCATAATGAGATATTATTA	66122
QY 1326	TCTGCAAGGTTTTTTTGTTGTGTGTTTTTTGTTTAAATTTTCAATATGCAAGT	1376
Db 66121	TCTGCAAGGTTTTTTTGTTGTGTGTGTTTTTTGTTTAAATTTTCAATATGCAAGT	66071
RESULT 8		
LOCUS	HSAL84ZAS	332 bp DNA STS 23-MAR-1996
DEFINITION	H.sapiens (D3S3570) DNA segment containing (CA) repeat; clone	
ACCESSION	AF041842s5; single read, sequence tagged site.	
VERSION	252396	
KEYWORDS	STS; CA repeat; dinucleotide repeat; GT repeat; microsatellite	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
	Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	1 (bases 1 to 332)	
AUTHORS	Weissenbach,J.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (01-SEP-1995) Genethon, B.P. 60, 91002 Evry Cedex France.	
AUTHORS	E-mail: Jean.Weissenbach@genethon.fr	
	2 (bases 1 to 332)	
	Dib,C., Faure,S., Fizanes,C., Samson,D., Drouot,N., Vignal,A.,	
	Mollasseau,P., Marc,S., Hazan,J., Seboun,E., Iathrop,M., Gyapay,G.,	
	Mollasseau,J. and Weissenbach,J.	
TITLE	A comprehensive genetic map of the human genome based on 5,264	
JOURNAL	microsatellites	
MEDLINE	Nature 380 (6570), 152-154 (1996)	
COMMENT	96176476	
FEATURES	full automatic.	
SOURCE	Location/Qualifiers	
	1. .332	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="3"	
	/cell_line="CEPH 134702"	
	/clone_id="genomic DNA"	
	/note="cloning vector is M13mp18"	
BASE COUNT	98 a	88 c 44 g 90 t 12 others
ORIGIN		
Query Match	8.2%;	Score 113; DB 13; Length 332;
Best Local Similarity	95.0%;	Pred. No. 2.8e-16;
Matches 113;	Conservative 0;	Mismatches 6; Indels 0; Gaps 0;
QY 1093	TGGAGAAAGGGAAGGCTCCCAACTTTGACACACACAAATATCAAGGCTCTTTGATA	1152
Db 327	TGGAGAAAGGGAAGGCTCCCAACTTTGACACACACAAATATCAAGGCTCTTTGATA	268

OY	1153	ATCAGTTTGATGGGATTTCACAAAGAAGAGTAAACAGAGGAAGCAAGAGTAGT	1211
Db	267	ATCACCTTTGATGTGGATTTCACANAGANGATTAAACAGAGGANGCAGAGANGTAGT	209
RESULT	9		
LOCUS	AF061749	2656 bp	mRNA PRT 08-AUG-1998
DEFINITION	Homo sapiens tumorous imaginal discs protein Tid56 homolog (TID1)		
ACCESSION	AF061749		
VERSION	AF061749.1	GI:3372676	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 2656) Schilling,B., De-Medina,T., Syken,J., Vidal,M. and Munger,K. A novel human DnaJ protein, Tld1-1, a homolog of the drosophila tumor suppressor protein Tid56, can interact with the human papillomavirus type 16 E7 oncoprotein Virology 247 (1), 74-85 (1998) 98354343		
JOURNAL	2 (bases 1 to 2656) Schilling,B., De-Medina,T., Syken,J., Vidal,M. and Munger,K. Direct Submission Submitted (27-APR-1998) Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115-5701, USA		
MEDLINE	Location/Qualifiers		
REFERENCES	1..2656		
AUTHORS	/organism="Homo sapiens"		
TITLE	/db_xref="taxon:9606"		
JOURNAL	/chromosome="16"		
FEATURES	/map="16p13.3: between D16S3070 and D16S510"		
source	1..2656		
gene	1..2656		
CDS	32..1474		
	/gene="TID1"		
	/note="Tid1"		
	/product="tumorous imaginal discs protein Tid56 homolog"		
	/protein_id="AAC29066.1"		
	/db_xref="GI:3372677"		
	/translation="MAARCSRLILLYVGCPRIAPISGRGARPPREGVAMSRKLS VPARSLTSQGPALPLTRPGVSLITGKNHPICLTAHFHSIAPLAEDRYOIIIGVRP NASQKEIKRAYIOLAKRRYPDNDDEPKAEFSQLAEAYEVLSDVKRKQDYAGSA GFDPGASGHSHYWGKPATVDPELPFKIFGEFSSSGDFQVFDDPOEFMELTFEN QAAQGVNKEFTVNIIMDCERNGKRPETVOHCHYGSGSMETINTGPFVMSTCR OCAGRGSIITSPCVCGAGGAKOKKKRMVIPAGVGDGOTVRMPVGRKFELTFRVQ KSYPFRPDGADIHSDLPISTIAQLLGSTARAQGLERTINVTPTGOTDKIRMGKG IPRNISTGYGDHTTHIKIRPVKRLTKRSQSLITSYAEDVDVEGTAVNGVITLSSGSST MDSSAGSKAREADEEGFLSKTKKMFYS"		
BASE COUNT	638 a 694 c 745 g 579 t		
ORIGIN			
Query Match	7.4% Score 101.4: DB 11; Length 2656;		
Best Local Similarity	61.6%; Pred. No.1,6e-13;		
Matches 162; Conservative	0; Mismatches 101; Indels 0; Gaps 0;		
OY	238	TACCATCATCGGGGGCGGTATTCGCCGAGACAGATTCCTATAAGATCTTGSGGGTCCCTCGA	297
Db	278	TTCCACACGAGTGCCTTTGGCCAAAGAAAGATTATTATCAGATTTAGAGAGTGCCTCGA	337
OY	298	AGTGCCTCTATAAAGATATTTAAAAAGGCTATAGAAAATAGACCCCTCAGCTTCATCCG	357
Db	338	AATGCCACAGCAAGAAAGAGATCAAGAAAGCTATATTATCAGCTTGCCCAAGATACACCT	397
OY	358	GACCGGAACCTCGATGATCCACAAGCCAGGAGAAATTCAGAGATCTGGGTCTCTTAT	417
Db	398	GACACAAATTAAGATGATCCCAAAGCCAGGAGGAGATTTCTCCCACTGGCAGAAAGCTAT	457

RESULT 12
LOCUS AB014888 1485 bp mRNA PRI 22-JAN-1999
DEFINITION Homo sapiens mRNA for MRJ, complete cds.
ACCESSION AB014888
VERSION AB014888.1 GI:3402484
KEYWORDS MRJ.
SOURCE Homo sapiens testis CDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 1485)
TITLE Saito,T. and Seki,N.
JOURNAL Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases.
Toshiyuki Saito, National Institute of Radiological Sciences,
Genome Research Group; Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
(E-mail:t.saito@nirs.go.jp, Tel:81-43-206-3135, Fax:81-43-251-9818)
2 (bases 1 to 1485)
Saito,T. and Seki,N.
REFERENCE A new member of human dnaJ-related gene family
AUTHORS Published Only in Database (1998) In press
JOURNAL Location/Qualifiers
FEATURES
source
1..1485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/tissue_type="testis"
101..826
/gene="MRJ"
101..826
/gene="MRJ"
/codon_start=1
/product="MRJ"
/protein_id="BAA32209.1"
/db_xref="GI:3402485"
/translation="WDYEVYGVORHASPEDIKKRYRLALKMHPDKNPENKEAE
KRYVAEAEVLSDAKRDITDKYKEGKNGGGSGSHDPSPEFTFRNDVRE
FTGGRDPSDFEDFEDFEDFEGNRKRGSRSGTSFSAFSGFSGFSGFSG
FTSGSLGHLGTLFSSTSGSGSGMGNFKSISTSTMVNGRKITTKRIVENGOERVE
EEDGOLKSLTLNKEQLRLDNK"
polya_signal 1459..1464
BASE COUNT 446 a 258 c 334 g 447 t
ORIGIN

Query Match 6.6%; Score 91.4; DB 9; Length 1485;
Best Local Similarity 66.5%; Pred. No. 3.3e-11;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATAGATCTTGGGGTGCCTCGAAGTCCCTATTAAGATATTAAAGGCC 327
|||||
DB 107 GATTACTATGAGTTCTAGCGCTGACAGACATGCTCACCGAGATTTAAAGCA 166
QY 328 TATAGGAAATACCCCTGCAGCTTCATCCGACCGAACCCTG--ATGATCCACAAGCC 384
|||||
DB 167 TATCGGAAACTGGCAGTCAAGTGCATCAAGATAAATATCTGAGATTAAGAAGCA 226
QY 385 CAGGAGAAATTCAGATCTGGGTGCTGCTTATGAGGTTCTGCATAGTAGAGAAACGG 444
|||||
DB 227 GAGAGAAATTTCAAGCAAGTAGCGAGGCGATATGAAGTGTGTCGATAGAAACGG 286
QY 445 AAACAGTACGATCTTATGTTGTAAGAAGATTAAGATG 485
|||||
DB 287 GACATCTATGACAAATATGCAAGAAGATTAATGTGG 327

RESULT 13
AB015799 1487 bp mRNA PRI 06-JAN-2000
LOCUS AB015799
DEFINITION Homo sapiens HSJ2 mRNA for DnaJ homolog, complete cds.
ACCESSION AB015799
VERSION AB015799.1 GI:6681593

KEYWORDS DnaJ homolog.
SOURCE Homo sapiens CDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 1487)
TITLE Hanai,R.
JOURNAL DnaJ homologue HSJ2b
JOURNAL Published Only in Database (2000) In press
AUTHORS 2 (bases 1 to 1487)
TITLE Hanai,R.
JOURNAL Submitted (24-JUN-1998) to the DDBJ/EMBL/GenBank databases. Ryo
Hanai, Rikkyo (St. Paul's) University, Department of Chemistry,
Nishi-Ikebukuro 3-34-1, Toshima, Tokyo 171-8501, Japan
(E-mail:hanai@rikkyo.ac.jp, Tel:81-3-3985-2377)
1..1487
Location/Qualifiers
FEATURES
source
1..1487
/organism="Homo sapiens"
/db_xref="taxon:9606"
148..873
/gene="HSJ2"
148..873
/gene="HSJ2"
/note="DnaJ homologue HSJ2b"
/codon_start=1
/product="DnaJ homolog"
/protein_id="BAA8770.1"
/db_xref="GI:6681594"
/translation="WDYEVYGVORHASPEDIKKRYRLALKMHPDKNPENKEAE
KRYVAEAEVLSDAKRDITDKYKEGKNGGGSGSHDPSPEFTFRNDVRE
FTGGRDPSDFEDFEDFEDFEGNRKRGSRSGTSFSAFSGFSGFSGFSG
FTSGSLGHLGTLFSSTSGSGSGMGNFKSISTSTMVNGRKITTKRIVENGOERVE
EEDGOLKSLTLNKEQLRLDNK"
BASE COUNT 436 a 270 c 343 g 438 t
ORIGIN

Query Match 6.6%; Score 91.4; DB 9; Length 1487;
Best Local Similarity 66.5%; Pred. No. 3.3e-11;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATAGATCTTGGGGTGCCTCGAAGTCCCTATTAAGATATTAAAGGCC 327
|||||
DB 154 GATTACTATGAGTTCTAGCGCTGACAGACATGCTCACCGAGATTTAAAGCA 213
QY 328 TATAGGAAATACCCCTGCAGCTTCATCCGACCGAACCCTG--ATGATCCACAAGCC 384
|||||
DB 214 TATCGGAAACTGGCAGTCAAGTGCATCAAGATAAATATCTGAGATTAAGAAGCA 273
QY 385 CAGGAGAAATTCAGATCTGGGTGCTGCTTATGAGGTTCTGCATAGTAGAGAAACGG 444
|||||
DB 274 GAGAGAAATTTCAAGCAAGTAGCGAGGCGATATGAAGTGTGTCGATAGAAACGG 333
QY 445 AAACAGTACGATCTTATGTTGTAAGAAGATTAAGATG 485
|||||
DB 334 GACATCTATGACAAATATGCAAGAAGATTAATGTGG 374

RESULT 14
AF075601 1489 bp mRNA PRI 11-JUL-1999
LOCUS AF075601
DEFINITION Homo sapiens heat shock J2 protein (HSJ2) mRNA, complete cds.
ACCESSION AF075601
VERSION AF075601.1 GI:5441949
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 1489)
TITLE Zhang,W., Wen,T., Yuan,Z., and Cao,X.
HSJ2, a novel human homologue of the bacterial heat-shock protein

Journal Reference Authors Title Journal
Dnaj Unpublished
2 (bases 1 to 1495)
Zhang, W., Wan, T., Yuan, Z. and Cao, X
Direct Submission
Submitted (01-JUL-1998) Department C

```
source      1..1489
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
gene        1..1489
            /gene="HSJ2"
CDS         133..858
            /gene="HSJ2"
```

BASE COUNT	443 a	269 c	337 g	440 t
ORIGIN				

Query Match	6.68;	Score 91.4;	DB 40;	Length 1489;
Best Local Similarity	66.5%;	Pred. No. 3.3e-11;		
Matches 147; Conservative	0;	Mismatches 71;	Indels 3;	Gaps 1;

RESULT	15
AF060703	
LOCUS	AF060703 1557 bp mRNA PRI 31-DEC-1999
DEFINITION	Homo sapiens DNAJ homolog mRNA, complete cds.
ACCESSION	AF060703
VERSION	AF060703.1 GI:6648622

```

/chromosome="11"
/cell_type="pancreatic cancer cells"
/map="11q"
146..871
CDS

```

BASE COUNT	472 a	275 c	357 g	452 t	1 others
ORIGIN					

Query Match	6.6%	Score 91.4;	DB 40;	Length 1557;
Best Local Similarity	66.5%;	Pred. No. 3.3e-11;		
Matches 147; Conservative	0;	Mismatches 71;	Indels 3;	Gaps 1;

Search completed: June 30, 2000, 12:46:48
Job time: 10107 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 12:12:13 ; Search time 929.99 Seconds
(without alignments)
5796.607 Million cell updates/sec

Title: US-09-501-714-4
Perfect score: 1330
Sequence: 1 CGNAGGAGAGAAAGAAAG.....TGACGCGACGGGTGCGGGG 1330

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

RESULT	1				
LOCUS	AL038833				
DEFINITION	AL038833	681 bp	mRNA	EST	27-SEP-1995
ACCESSION	DKEF2566C00946	_t1	566 (synonym: hfk42)	Homo sapiens	CDNA clone
VERSION	DKEF2566C00946	5'			mRNA sequence.
KEYWORDS	AL038833				
SOURCE	AL038833.1	GI:5407962			
ORGANISM	EST.				
	human.				
	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
AUTHORS	1 (bases 1 to 681)				
	Ottewaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and				
	Wiemann, S.				
TITLE	EST (Ottewaelder, et al.)				
JOURNAL	Unpublished (1995)				
COMMENT	On Mar 10, 1998 this sequence version replaced gi:2948196.				
	Contact: Ottewaelder B				
	MIPS				
	Am Klopfersplitz 18a D-82152 Martinsried, Germany				

This is the 5' sequence of the clone inset from S. Wiemann, Molecular Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MedGenomix within the cDNA sequencing consortium of the German Genome Project. si sequence also available.

This clone is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES	Location/Qualifiers
source	1. .681

```

/organism="Homo sapiens"
/ab_xref="taxon:9606"
/clone="DKFZP56600946"
/clone_lib="566 (synonym: hfk42)"
/clisse_type="kidney"
/dev_stage="fetal"
/lab_host="xl-2blue"
/notes="vector: pMP1, Site_1: NotI, Site_2: SalI"
BASE COUNT      177 a      151 c      163 g      170 t
ORIGIN

```

Query Match	45.38;	Score 616;	DB 61;	Length 681;
Best Local Similarity	97.08;	Pred. No. 1.4e-149;		
Matches 644; Conservative	0;	Mismatches 2;	Indels 2;	Gaps 2;

QY	23	GC	CGCAGAGAGCCGCCCGCNACCACACAGCGNCA	CANTCTGTGNC	GTN	AGAGATTCGGGC	82
Db	11	GC	CGAGAGAGCGCGCGCCACACACAGGCGCAG	CACTCTCTG	AGAGATTCGGGC	70	
QY	83	CGTCA	CCCTGCGCTCCGCTGCTCCCGCACCGGGCG	CTTCTTTC	CTGTGAGCCATTCCAA	142	
Db	71	CGTCA	CCCTGCGCTCCGCTGCGCACCGGGCGCTT	CTGCTCT	GTGAGCCATTCCAA	130	
QY	143	CAATCT	CGTAAACAATG	TGATTA	CTAAGTTCTT	AGCGCTGCAGACATGCTCAC	202
Db	131	CAATCT	CGTAAACAATG	TGATTA	CTAAGTTCTT	AGCGCTGCAGACATGCTCAC	190
QY	203	CCGAGATAT	TAATAAAAAGCATATGCGAAATGCGAC	ACTGAGTGGCATTCG	CAATGATTAATC	262	
Db	191	CCGAGATAT	TAATAAAAAGCATATGCGAAATGCGAC	ACTGAGTGGCATTCG	CAATGATTAATC	250	
QY	263	CTGAGAT	TAATAAGAGAGAGCAGAGAAATTC	CAAGCA	GTAGCGGAGCATATGAAGTGC	322	
Db	251	CTGAGAT	TAATAAGAGAGAGCAGAGAAATTC	CAAGCA	GTAGCGGAGCATATGAAGTGC	310	
QY	323	TGTGCGAT	TGCTAAGAAACGGGACATCATGACAAAT	ATATGGCA	AGAAAGATTAAATGGTG	382	
Db	311	TGTGCGAT	TGCTAAGAAACGGGACATCATGACAAAT	ATATGGCA	AGAAAGATTAAATGGTG	370	
QY	383	GNGNGNG	NGGTGGAAAGCATTTTGGACAGTCCATTT	GAAATTTGGCTTC	CACATTCGGTAAC	442	
Db	371	GAGGAGG	AGGTGGAAAGCATTTTGGACAGTCCATTT	GAATTTGGCTTC	CACATTCGGTAAC	430	
QY	443	CAGATGAT	GTCTTCAGGGAAATTTTGGTGGAAAGGAG	ACCATTCTT	CATTTGACTTCCTTG	502	
Db	431	CAGATGAT	GTCTTCAGGGAAATTTTGGTGGAAAGGAG	ACCATTCTT	CATTTGACTTCCTTG	490	
QY	503	AAGAC	CCCTTTGAGGACTTCTTTGGGAATCGAAGGG	GGTCCC	CGAGAGACAGAGCCGAG	562	
Db	491	AAGAC	CCCTTTTGGAGCACTTCTTTGGGAATCGAAG	GGGGTCCC	CGAGAGACAGAGCCGAG	550	
QY	563	GGA	CGGGGTCGTTTTCTCTGCGTCA	GTAGGATTC	CCGCTTTTGGAAATGATTTCTG	622	
Db	551	GGA	CGGGGTCGTTTTCTCTGCGTCA	GTAGGATTC	CCGCTTTTGGAAATGATTTCTG	609	
QY	623	CTTTTA	TATACAGATTACTCTATTTGGGT	CAC	TAGAGTACG	GGGGGCGCTCACTTCATTC	682
Db	610	CTTTTA	TATACAGATTACTCTATTTGGGT	CAC	TAGAGTAC	GGGGGCGCTCACTTCATTC	668
QY	683	CTTC	686				

Am Klopferpfütz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ): Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Medicinomics within the CDNA sequencing consortium of
the German Genome Project.
sl sequence also available.
This clone is available at the RZPD in Berlin.
Please contact the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

1. .661

Query Match	43.5%	Score 578.8	DB 61	Length 661
Best Local Similarity	95.7%	Pred. No. 6.3e-140		
Matches 622; Conservative	0	Mismatches 25	Indels 3	Gaps 3

RESULT	3	AA020916	LOCUS	DEFINITION
		AA020916	646 bp	mRNA EST 30-JAN-1997
		zee44h08.r1		Soares retina N2b4HR Homo sapiens cDNA clone
		IMAGE:563807.5		similar to PIR:S25509 S25509 dnaJ protein homolog
		human ;		mRNA sequence.

ACCESSION	AA020916
VERSION	AA020916.1
KEYWORDS	EST.
SOURCE	human.

ORGANISM Homo sapiens

Eukaryōta; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE	AUTHORS
1 (bases 1 to 646)	Hallier, L., Lemon, C., Becker, M., Ronaldo, M. F., Chiapelli, B., Chissoc, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Harkins, M., Hultman, M., Kucab, T., Lacy, M., Le, N., Maddis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rflkin, L., Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevaslis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, M. and Marra, M.
TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
COMMENT	97044478
	On May 8, 1995 this sequence version replaced gi:79404.

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAG Consortium (info@image.llnl.gov) for further information.
Insert length: 1524 Std Error: 0.00
Seq primer: *28M13 rev2 from Amersham
High quality sequence stop: 478.

FEATURES	Location/Qualifiers
source	1. .646

```

/organism="Homo sapiens"
/db_xref="GDB:1280510"
/db_xref="taxon:9606"
/clone_image="363807"
/clone_11b="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pRT33 (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - o19o(OT) primer [5

```


|||||
Db 359 GGCTTCATTCGGAACCCAGATGATGTCTTCAGGGAATTTTGGTGAAGGACCA 418
OY 484 TTTTCATTGACTCTCTTTGAAGACCCCTTTGAGGACTCTTTGGGAATCGAAGGGCTCC 543
Db 419 TTTTCATTGACTCTCTTTGAAGACCCCTTTGAGGACTCTTTGGGAATCGAAGGGCTCC 478
OY 544 CGAGAACCAAGACCGGAGCGGCTCTTTCTGCTGCTTACGTGATTCCTGCT 603
Db 479 CGAGAACCAAGACCGGAGCGGCTCTTTCTGCTGCTTACGTGATTCCTGCT 538
OY 604 TTTGGAAGTGGATT 618
Db 539 TTTGGAAGTGGATT 553
RESULT 5
AA993699/c
LOCUS
DEFINITION AA993699 529 bp mRNA EST 27-AUG-1998
OC97C10.s1 Soares_total_fetus_ND2HF8_9w Homo sapiens cDNA clone
IMAGE:1624722 3' similar to TR:035723 035723 TESTIS SPECIFIC
DNAJ-HOMOLOG.; contains element MSRI MSRI repetitive element ;,
mRNA sequence.
ACCESSION AA993699 GI:3180244
VERSION AA993699.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 529)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 700 Std Error: 0.00
Seq primer: -40ml3 fwd. Ex from Amersham
High quality sequence stop: 457.
Location/Qualifiers
FEATURES
Source
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1624722"
/clone_lib="Soares_total_fetus_ND2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
GTTACCAATCGAAGTGGAGCGCGGCGTAAATTTTCTTTTCTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 145 a 143 c 89 g 152 t
ORIGIN
Query Match 38.1%; Score 506.6; DB 40; Length 529;
Best Local Similarity 97.7%; Pred. No.3.4e-121;
Matches 512; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
OY 157 ATGTGATTAATGAAATCTAGCGCTGCAGACATGCTCACCGAGGATTTAA 216
|||||
Db 529 ATGTGATTAATGAAATCTAGCGCTGCAGACATGCTCACCGAGGATTTAA 470

OY 217 AAGCATATCGGAACCTGGACATGAGTGCATCCAGATTAATAATCTGAGATAAAGA 276
Db 469 AAGCATATCGGAACCTGGACATGAGTGCATCCAGATTAATAATCTGAGATAAAGA 410
OY 277 GAAGCAGAGAGAAATTCAGCAAGCAATAGCGAGGATATGAGTCTGCGATCTAAG 336
Db 409 GAAGCAGAGAGAAATTCAGCAAGCAATAGCGAGGATATGAGTCTGCGATCTAAG 350
OY 337 AAACGGACATCTATGACAAATATGCAAGAAGATTAATATGATGAGNGAGNGGTGA 396
Db 349 AAACGGACATCTATGACAAATATGCAAGAAGATTAATATGATGAGNGAGNGGTGA 290
OY 397 AGTCATTTTGGACATGCAATTTTGAATTTGGCTTCAATTCGCCAATCCAGATGATCTTC 456
Db 289 AGTCATTTTGGACATGCAATTTTGAATTTGGCTTCAATTCGCCAATCCAGATGATCTTC 230
OY 457 AAGCAATTTTGGTGGGAAGGAGGACCATTTTCATTTGACTTCTTGAAGACCTTTTGA 516
Db 229 AAGCAATTTTGGTGGGAAGGAGGACCATTTTCATTTGACTTCTTGAAGACCTTTTGA 170
OY 517 GACTTCTTTGGGAATCGAAGGGGTCGCCGAGAGACAGAACCGGAGGAGGCTGTT 576
Db 169 GACTTCTTTGGGAATCGAATTTTTCGCCGAGAGACAGAACCGGAGGAGGCTGTT 110
OY 577 TTTCTGCGCTTACGTGATTTTCCTTTTGAATGATTTTCTTTTATGATACAGA 636
Db 109 TTTCTGCGCTTACGTGATTTTCCTTTTGAATGATTTTCTTTTATGATACAGA 50
OY 637 TTTACTCATTTGGGTCATAGTACAGGGGAGCTCATTCAT 680
Db 49 TTTACTCATTTGGGTCATAGTACAGGGGAGCTCATTCAT 6
RESULT 6
AI028519/c
LOCUS
DEFINITION AI028519 535 bp mRNA EST 28-AUG-1998
OM44b09.x1 Soares_parathyroid_tumor_NbHRA Homo sapiens cDNA clone
IMAGE:1649657 3' similar to TR:035723 035723 TESTIS SPECIFIC
DNAJ-HOMOLOG. ; mRNA sequence.
ACCESSION AI028519
VERSION AI028519.1 GI:3245828
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 535)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/dbtrp/image/image.html
Insert Length: 668 Std Error: 0.00
Seq primer: -40ml3 fwd. Ex from Amersham
High quality sequence stop: 450.
Location/Qualifiers
FEATURES
Source
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1649657"
/clone_lib="Soares_parathyroid_tumor_NbHRA"
/tissue_type="parathyroid tumor"

BASE COUNT	179 a	147 c	175 g	152 t	8 others
ORIGIN					
Query Match					36.7%; Score 488; DB 26; Length 661;
Best Local Similarity					93.6%; Pred. No. 2,5e-116;
Matches 559;					Conservative 0; Mismatches 31; Indels 7; Gaps 5;
OY	23	GCCCGAGAGACCCCGCCGACACACAGCGNCACANTCTCGG-NGCTNTGAGAGATTCGGG	81		
Db	14	GCGGAGAGAGCGCGCCGACACACAGCCACAGNCCTCGANNNTGTAGAGGATTCGGG	73		
OY	82	CCGTGACCCCTGCTCCCGCTCCGCTCCGCGACCGCGCTCTTTCTCTGGAGCCCAATTCGA	141		
Db	74	CCGTACCCCTGCTCCCGCTCCGCGACCGCGCTCTTTCTCTGGAGCCCAATTCGA	133		
OY	142	ACAATCTCGTAAACATGATGGTATCTATGAACTTCTAGCGGTCGACAGACATGCTCA	201		
Db	134	ACAATCTCGTAAACATGATGGTATCTATGAACTTCTAGCGGTCGACAGACATGCTCA	193		
OY	202	CCCGAGATATTTAAAGGACATATCGGAACTGCGACTGGAAGTGGCATCCAGATATAAAT	261		
Db	194	CCCGAGATATTTAAAGGACATATCGGAACTGCGACTGGAAGTGGCATCCAGATATAAAT	253		
OY	262	CCTGAGATTAAGAAAGAGAGAGAAATTCAGCAAGATAGCGGAGGACATATGAATG	321		
LOCUS	254	CCTGAGATTAAGAAAGAGAGAGAAATTCAGCAAGATAGCGGAGGACATATGAATG	313		
OY	322	CTGTGCGATGCTAAAGAAAGCGGACATCTATGCAATATATGCGAAAGAGATTAATG	381		
Db	314	CTGTGCGATGCTAAAGAAAGCGGACATCTATGCAATATATGCGAAAGAGATTAATG	373		
OY	382	GGNGGNGNGTGTGAAGTCATTTTGACAGTCCATTTGAATTTGGCTTCACATTCGGTAC	441		
Db	374	GGAGAGAGAGGNGGAAATTCATTTTGACAGTCCATTTGAATTTGGCTTCACATTCGGTAC	433		
OY	442	CCAGATGATGCTTTCAGGGAATTTTGGTGGAAAGGACCATTTTTCATTTGACTTCTT	501		
Db	434	CCAGATGATGCTTTCAGGGAATTTTGGTGGAAAGGACCATTTTTCATTTGACTTCTT	493		
OY	502	GAAAGACCTTTTGAAGACTCTTTGGGATGCAAGAGGGTCCCGAGAGAGCAAGACCGA	561		
Db	494	GAAAGACCTTTTGAAG-ACTCTTTGGGATGCAAGAGGCT-CCCGAGAGAGCAAGACCGA	550		
OY	562	GGGAGGGGGTCTTTTCTCTGCGTTCAGTGGATTTCCGTTTGGAGATGGATTT	618		
Db	551	AGGAGCGGGG-CGNTTTTCCCTGCTTCAGTGGATTT-CCGCTTGGAGATGGATTT	604		
RESULT	9				
AL036026					
LOCUS	AL036026	531 bp	mRNA	EST	27-SEP-1999
DEFINITION	DKFZP564P0422.r1.564 (synonym: h6br2) Homo sapiens cDNA clone				
ACCESSION	AL036026				
VERSION	AL036026.1	GI:5405652			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	Manbuti,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.				

OY	443	CAGATGAGTCCTTGGAGGAATTTTTGCGAAGGACCACTTTCATTGACTCCTTG	502
Db	431		490
OY	503	AAGACCCTTTTGAGACTTC	522
Db	491		510
		AAGACCCTTTTGAGACTTC	510
RESULT	11		
Locus	AW239553		
DEFINITION	AM239553	545 bp mRNA EST	13-DEC-1999
ACCESSION	x632g09.y1 NCI-CGAP_Lu31 Homo sapiens cDNA clone IMAGE:2578048 5'		
VERSION	Similar to TR:095806 095806 DNA-Like 2 PROTEIN. ; mRNA sequence.		
KEYWORDS	AW239553		
SOURCE	EST.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homimidae; Homo.		
TITLE	1 (bases 1 to 545)		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
	Unpublished (1997)		
	On May 18, 1998 this sequence version replaced gi:3138203.		
	Other ESTs: x632g09.x1		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: ATCC CDNA Library Preparation: Life		
	Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The		
	I.M.A.G.E. Consortium DNA Sequencing by: Washington University		
	Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at:		
	www-bio.lnl.gov/db/rp/image/image.html		
FEATURES			
Source	Seq primer: -40RP from Gibco		
	High quality sequence stop: 373.		
	Location/qualifiers		
	1..545		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2578048"		
	/clone_lib="NCI-CGAP_Lu31"		
	/sex="male"		
	/dev_stage="fetal, 14 wk post-conception"		
	/lab_host="DH10B"		
	/note="Organ: lung, cell line: Vector: pCMV-SPORT6; site:1: EcoRV; Site:2: NotI; Cloned unidirectionally, no 5' adaptor. Primer: Oligo dr. Full-length library constructed by Life Technologies."		
BASE COUNT	155 a	141 g	131 t
ORIGIN			1 others
Query Match	35.7%; Score 474.8; DB 79; Length 545;		
Best Local Similarity	94.2%; Pred. No. 6.3e-113;		
Matches 502; Conservative	0; Mismatches 30; Indels 1; Gaps 1;		
OY	74	GATTGGGCGGTGACCCCTGCCCTGCCTGCCGCCACCGGCGCTTCTTCTCGGAC	133
Db	1		60
	GATTGGGCGGTGACCCCTGCCCTGCCTGCCGCCACCGGCGCTTCTGCTCGGAC		
OY	134	CCATTCCAACAATCTGTAATAAACATGATGATGATGATGAACTTCTAGCGCTGACAGAC	193
Db	61	CCATTCCAACAATCTGTAATAAACATGATGATGATGATGAACTTCTAGCGCTGACAGAC	120
OY	194	ATGGCTCACCCGAGATATTAAAAAGCATATCGGAAACTGGCATGAAGTGGCATCCAG	253
Db	121	ATGGCTCACCCGAGATATTAAAAAGCATATCGGAAACTGGCATGAAGTGGCATCCAG	180

OY	254	ATAAAAATCCGAAATAAAGGAAGCAGAGAAGAAATTCAACGAAGTGGCGAGCAT	313
Db	181	ATAAAAATCTTGAGACTAAAGAAAGCACAGAGAAAATTTCAAGCAAGTAGCGGAGCAT	240
OY	314	ATGAAGTCGTCTGCAGTCTTAAGAAACGGGACATCTATGACAATAATGGCAAGAAGAT	373
Db	241	ATGAAGTGTGTGTGCGATGCTAAGAAACGGGACATCTATGACAATAATGGCAAGAAGAT	300
OY	374	TAAATGCTGNGNGNGNGTGGAAATCATTTTGCAGATGCCATTTGAATTTGGCTTCACAT	433
Db	301	TAAATGCTGNGNGNGNGTGGAAATCATTTTGCAGATGCCATTTGAATTTGGCTTCACAT	360
OY	434	TCCGTAACCCAGATGATGCTTCACGGAATTTTTTGGTGAAGGGACCACAT-TTTCATTT	492
Db	361	TCCGTAACCCAGATGATGCTTCACGGAATTTTGTGTGAAGGGACCACATTTTCATTT	420
OY	493	GACTTCTTTTGAAGACCCCTTTTGAAGACTTCTTTGGGAATCGAAGGGGTCCCAGAGAAGC	552
Db	421	GACTTCTTTTGAAGACCCCTTTTGAAGACTTCTTTTGAAGATCGAAGGGGTCCCAGAGAAGC	480
OY	553	AGAACCCGAGGAGACGGGGTCTTTTCTCTGCGCTTACGTGATTTTCGGCTTTT	605
Db	481	TAAAGCCGAGGAGACGGGGTCTTTTCTCTGCGCTTACGTGATTTTCGGCTTTT	533
RESULT 12			
LOCUS	AU080965	804 bp mRNA EST	20-OCT-1999
DEFINITION	AU080965 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-6291		
ACCESSION	AU080965	5', mRNA sequence.	
VERSION	AU080965.1	GI:6085719	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (Phases 1 to 804)		
JOURNAL	Hashimoto,K., Kusuda,J., Toyoda,A., Tanuma,R., Ito,A., Hirata,M.,		
COMMENT	Suzuki,Y., Sasaki,M. and Sugano,S. Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method Unpublished (1999) On Jun 5, 1998 this sequence version replaced gi:3189734. Contact: Katsuyuki Hashimoto Division of Genetic Resources National Institute of Infectious Diseases 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan Email: khashie@nih.go.jp URL: http://www.nih.go.jp/tokyo/genbank/ location/Qualifiers		
FEATURES	source		
	1..804		
	/organism="Mus musculus"		
	/strain="C57BL"		
	/db_xref="taxon:10090"		
	/clone="MNCB-6291"		
	/clone_lib="Sugano mouse brain mncb"		
	/sex="female"		
	/dev_stage="adult"		
	/lab_host="TOP10"		
	/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT] double-stranded cDNA was ligated to a DraIII adaptor [TTGTGACCTACGG], digested and cloned into distinct DraIII sites of the pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed by exclude fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer for sequencing: 5' end primer [CTTTCGTCTTAAAGCTGCG]		
BASE COUNT	226 a 161 c 205 g 199 t		13 others

LOCUS	AA069940	486 bp	mRNA	EST	23-DEC-1997
DEFINITION	zm59a07.r1 Stratiogene fibroblast (#937212) Homo sapiens cDNA clone IMAGE:529908 5' similar to SW-HS1_HUMAN P25686 DNAJ PROTEIN				
ACCESSION	AA069940				
VERSION	AA069940.1	GI:1577301			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 486) Hillier,L., Lemon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissee,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mandis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry,Meg,J., Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,M. and Marra,M.				
TITLE	Generation and analysis of 280,000 human expressed sequence tags				
JOURNAL	Genome Res. 6 (9), 807-828 (1996)				
MEDLINE	97044478				
COMMENT	On Apr 14, 1993 this sequence version replaced gi:692385. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu WARNING: There is evidence that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.				
FEATURES	<p>This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1304 Std Error: 0.00 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 356. Location/Qualifiers</p> <p>1..486 /organism="Homo sapiens" /db_xref="GDB:3919419" /db_xref="taxon:9606" /clone="IMAGE:529908" /clone_lib="Stratiogene fibroblast (#937212)" /lab_host="SOLR cells (kanamycin resistant)" /note="Vector: pBluescript SK-; site.1: EcoRI; site.2: XhoI; cloned unidirectionally. Primer: Oligo dT. M13 cell line. Average insert size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"</p>				
BASE COUNT	129 a	85 c	127 g	144 t	1 others
ORIGIN					
Query Match	34.6%	Score 460.4	DB 28	Length 486	
Best Local Similarity	98.5%	Pred. No. 3.3e-109			
Matches 472	Conservative 1	Mismatches 5	Indels 1	Gaps 1	
0Y	376	AATGCTGNGNGGNGGAGTGAAGTATTTTGGACAGTCACATTTGAAATTTGGCTTCACATTC	435		
Db	1	AATGGTGAGGAGGAGGAGGTGGAGTATTTTGGACAGTCACATTTGAAATTTGGCTTCACATTC	60		
0Y	436	CCTAACCCAGATGATCTTCACAGGAATTTTTTGGTGCAAGGACCCATTTTCATTTGAC	495		
Db	61	CCTAACCCAGATGATCTTCACAGGAATTTTTTGGTGCAAGGACCCATTTTCATTTGAC	120		
0Y	496	TTCTTTGAAGACCCCTTTTGAAGACTTCTTTGGGAATCGAAGGGGTCCCGAGGAAGCACA	555		
Db	121	TTCTTTGAAGACCCCTTTTGAAGACTTCTTTGGGAATCGAAGGGGTCCCGAGGAAGCACA	180		

OY	556	AGGAGGGACAGGGGGTC-GTTTTCTCGCGCTTAGAGAAATTCGCTTTTGAAGTG	61.4
Dd	181	AGCCGAGGGACGGGGTCNGTTTTCTCGCCTTAGAGATTCCGTCTTTGSAATGG	240
OY	615	ATTTCCTCTTTATACAGAGATTACTTCATTGGGGTCATAGTCACGGGGGCTTCAC	67.4
Dd	241	ATTTCCTCTTTTGATACAGSATTACTCAATTTGGGTCATAGTCACGGGGGCTTCAC	300
OY	675	TTTCATTCTCTCCACGCATTTTGGTGTAGTGGCATGGGCCAATTCAAATGATATCAAC	73.3
Dd	301	TTTCATTCTCTCCACGCATTTTGGTGTAGTGGCATGGGCCAATTCAAATGATATCAAC	360
OY	735	TTCACTTAATAATGGTTATAGGCAGAAAAATCACACAAGAAGATTTTCGAAACGGTCA	79.4
Dd	361	TTCACTTAATAATGGTTATAGGCAGAAAAATCACTACAAGAAGATTTTCGAAACGGTCA	420
OY	795	AGAAAGACTAGAACGTTGAGAAGATGGCCACTTAAGTCTCTTAACAATAATGGTGTGK	85.3
Dd	421	AGAAAGACTAGAACGTTGAGAAGATGGCCACTTAAGTCTCTTAACAATAATGGTGTAGG	479
RESULT	15		
	A1048617		
LOCUS			
DEFINITION			
	A1048617	655 bp mRNA	EST
		u64aa02.yl Sugano mouse liver.mlia Mus musculus cDNA clone	08-JUL-1998
		IMAGE:1450634.5' similar to TR:035723 O35723 TESTIS SPECIFIC	
ACCESSION		DNAJ-HOMOLOC.; , mRNA sequence.	
VERSION		A1048617	
KEYWORDS		A1048617.1 GI:3296904	
SOURCE		EST.	
ORGANISM		house mouse.	
		Mus musculus	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
		Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.	

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 655)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,B.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMMI Mouse EST Project
Unpublished (1996)
On Jan 14, 1998 this sequence version replaced gi:1798000.
Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine#
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNC+; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGF:923950
Seq primer: custom primer used
High quality sequence stop: 516.
Location/Qualifiers
1..655
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1450634"
/clone_lib="Sugano mouse liver ml1a"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: Liver; Vector: pME18S-FL3; Site_1: DraIII
(CACCTGTGTG); Site_2: DraIII (CACCACTGTG); 1st strand cDNA
was primed with an oligo(dT) primer
(AATGGCCCTTTTTTTTTTTTTT); double-stranded cDNA was
ligated to a DraIII adaptor (TGTGGCCCTGTG), digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACCTGTGTG, 3' site CACCACTGTG). XhoI should
be used to isolate the cDNA insert. Size selection was

```

performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTTAAAGCTGGC and 3' end
primer CGACCTGCAGCTCGACACA."

BASE COUNT 172 a 148 c 175 g 160 t
ORIGIN

Query Match 34.0%; Score 452; DB 41; Length 655;
Best Local Similarity 82.8%; Pred. No. 5.8e-107;
Matches 552; Conservative 0; Mismatches 99; Indels 16; Gaps 3;

```
QY 12 AAAGGAAAGNCGCGGAGAGCCGCGCMACACAGCGNCACANTCCTGGNGCTNTGAG 71
    || || || || || || || || || || || || || || || || || || || ||
Db 2 AAGCGGAGAGAGAGCGCGCGCGCCACACCGGCACAGGCGAGCTGTGCGACAGCGCGAG 61
    || || || || || || || || || || || || || || || || || || || ||
QY 72 GAGATTGGGCGCGTCACTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 131
    || || || || || || || || || || || || || || || || || || || ||
Db 62 GAGA--CGGCGCGCGCTGACTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 108
    || || || || || || || || || || || || || || || || || || || ||
QY 132 ACCCATTCCAACATCTCGTAAACATGTGGATTACTATGAAGTCTAGGCGTGCAGAG 191
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109 ATTATTCCAAACACTCAGTTAAACATGTGGATTACTATGAAGTCTAGGCGTGCAGAG 168
    || || || || || || || || || || || || || || || || || || || ||
QY 192 ACATGCTCCACCGGAGATATTTAAAGCATATCGGAACTGGCACTGAAGTGCATCC 251
    || || || || || || || || || || || || || || || || || || || ||
Db 169 ACATGCTCCACCGTGAAGATTTAAAGCGGTATCGAAACAGGCACTTAATGGCACCC 228
    || || || || || || || || || || || || || || || || || || || ||
QY 252 AGATTAATAATCCTGAGAAATAAGAAAGAGAGAAATTCAGCAAGTAGCGAGGC 311
    || || || || || || || || || || || || || || || || || || || ||
Db 229 GGACAAAAATCTGTAATAATAAGAAAGAGAGAGCGGAGTTCAACAGTAGCTGAGGC 288
    || || || || || || || || || || || || || || || || || || || ||
QY 312 ATATGAAGTCTGTGAGATCTAAGAAAGGACATCTATGACAAATATGCAAGAAGG 371
    || || || || || || || || || || || || || || || || || || || ||
Db 289 ATATGAAGTGTATGAGATCTAGAAAGCGGACATCTAGCACAATATGCAAGAAGG 348
    || || || || || || || || || || || || || || || || || || || ||
QY 372 ATTAAA---TGGTGGNGNGNGGTGGAAGTCTTTTGTGACAGTCCATTGGAATTTGGCTT 428
    || || || || || || || || || || || || || || || || || || || ||
Db 349 ATTAAATGTGTGAGAGAGAGGTGGAATTTTGTGACAGTCCATTGGAATTTGGCTT 408
    || || || || || || || || || || || || || || || || || || || ||
QY 429 CACATTCGCTAACCCAGATATGTCTCAGGAAATTTTGTGGAAGGAGCCATTTC 488
    || || || || || || || || || || || || || || || || || || || ||
Db 409 CACATTCGCGAACCAGATATGTCTCAGGAAATTTTGTGGAAGGAGCCATTTC 468
    || || || || || || || || || || || || || || || || || || || ||
QY 489 ATTGACTTCTTGAAGACCTTTTGAAGTCTTGTGGAATGGAAGGGTCCCGAGG 548
    || || || || || || || || || || || || || || || || || || || ||
Db 469 ATTGACTTCTTGAAGACCTTTTGAAGTCTTGTGGAATGGAAGGGTCCCGAGG 528
    || || || || || || || || || || || || || || || || || || || ||
QY 549 AAGCAGAAAGCGAGGAGCGGGTCTTTTCTGCGTTCAGTGATTTCCGTCCTTTGG 608
    || || || || || || || || || || || || || || || || || || || ||
Db 529 AAATAGAAAGCGAGGAGTCCGCGTCAATTTTCTGACCTCAGTGATTTCCCTTTTGG 588
    || || || || || || || || || || || || || || || || || || || ||
QY 609 AAGTGGATTCTTCTTTTATACAGATTTACTTCAATTTGGGTCACTAGGTACGGGGG 668
    || || || || || || || || || || || || || || || || || || || ||
Db 589 AAGTGGATTCTTCTTTTATACAGGCTTCACTCAATTTGGGTCACTAGGTATGGGGG 648
    || || || || || || || || || || || || || || || || || || || ||
QY 669 CTTCACT 675
    || || || || || || || || || || || || || || || || || || || ||
Db 649 TCTCACT 655
    || || || || || || || || || || || || || || || || || || || ||
```

Search completed: June 30, 2000, 12:12:25
Job time: 8076 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 09:57:49 ; Search time 929.99 Seconds

(without alignments)
5997.092 Million cell updates/sec

Title: US-09-501-714-2

Perfect score: 1376

Sequence: 1 TCTCAGCGGACTCGGACT.....TTTATTTCATATGCAACT 1376

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
1	681	49.5	735	69	AM11601	AM11601 x132a01.x
2	627.2	45.6	656	64	AM027159	AM027159 wt7207.x
3	619.4	45.0	695	36	AA639658	AA639658 n8307.s
4	572.8	41.6	655	40	AA984302	AA984302 am8301.s
5	484.8	35.2	498	36	AA599885	AA599885 ag32e11.s
6	481.2	35.0	488	35	AA519955	AA519955 n13605.s
7	476.8	34.7	482	25	N93316	N93316 zb69e03.s1
8	455.6	33.1	791	61	AI819362	AI819362 w61b11.x
9	445.4	32.4	767	80	AA837534	AA837534 oe31e05.s
10	441.1	32.0	556	80	AA333622	AA333622 u074e06.y
11	439.8	32.0	750	35	AA596749	AA596749 vm5e09.r
12	423.8	30.8	438	26	W65314	W65314 zd33e05.r1
13	417	30.3	728	60	AI800795	AI800795 w913e05.x
14	415.4	30.2	480	69	AA122551	AA122551 UT-M-BH2
15	406.6	29.5	487	34	AA512652	AA512652 v119f11.r
16	400	29.1	449	26	W90172	W90172 zh78a08.r1
17	399	29.0	498	30	AA204094	AA204094 m31e01.r
18	397.6	28.9	721	62	AI925702	AI925702 w033e11.x
19	395	28.7	423	26	W67505	W67505 zd40f10.r1
20	387.6	28.2	423	26	W67505	W67505 zd40f10.r1
21	387.2	28.1	426	31	AA654238	AA654238 zs91c06.s
22	387.2	28.1	426	31	AA654238	AA654238 zs91c06.s
23	387.2	28.1	426	31	AA654238	AA654238 zs91c06.s
24	375.4	27.3	683	51	AI742784	AI742784 w946b12.x
25	369.2	26.8	377	63	AI972560	AI972560 wr39b03.x
26	352	25.6	455	26	W56612	W56612 zd16a09.r1
27	344.4	25.0	359	48	C84272	C84272 C84272 oste
28	341.4	24.8	382	80	AA376471	AA376471 IL3-CT021
29	339.4	24.7	351	43	AA1202493	AA1202493 gs69n01.x
30	339	24.6	453	34	AA497706	AA497706 v168d03.r
31	336	24.0	378	74	AM207301	AM207301 UT-H-B11
32	330.4	24.0	346	36	AA639028	AA639028 ns02b03.s
33	315.8	23.0	631	80	AA634328	AA634328 f174h05.y
34	312.6	22.7	417	48	AI614268	AI614268 v168d03.y
35	312.6	22.7	451	24	N42272	N42272 yw85f08.r1
36	311.6	22.6	608	63	AI980687	AI980687 pat PK003
37	307	22.3	497	74	AI996659	AI996659 da09h04.y
38	305.6	22.2	449	22	R72988	R72988 y138e08.r1
39	305	22.1	371	29	AA144567	AA144567 mr64a04.r
40	304.6	22.1	370	36	AA607886	AA607886 vm40e04.r
41	304	22.1	327	32	AA339144	AA339144 ST44230
42	301.2	21.9	410	28	AA121115	AA121115 zm22h04.r
43	300.4	21.8	413	25	N86926	N86926 LL478F Homo
44	300	21.8	311	45	AI351856	AI351856 q903h09.x

ALIGNMENTS

RESULT 1
 LOCUS AM11601
 DEFINITION x132a01.x1 NCI-CGAP_Brn50 Homo sapiens cDNA clone IMAGE:2619720 3' similar to WP:115H9.1 CE0164 HYPOTHETICAL 39.9 KD PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR ; mRNA sequence.

ACCESSION AM11601
 VERSION AM11601.1
 KEYWORDS GI:6133208
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index
 Unpublished (1998)

JOURNAL

On Dec 20, 1995 this sequence version replaced gi:1135183.
 Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Burt Feuerstein, M.D., Mark Israel, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Possible reversed clone: similarity on wrong strand
 Seq primer: -40UP from Glibco
 High quality sequence stop: 471.

FEATURES

source

1. 735
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2619720"
 /clone_lib="NCI-CGAP_Brn50"
 /tissue_type="medulloblastoma"
 /lab_host="DH10B (phage resistant)"
 /note="Organ: brain; Vector: pTRT3-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from medulloblastoma tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTRT3 vector. This library is normalized. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 228 a 132 c 203 g 168 t 4 others
 ORIGIN

Query Match 49.5%; Score 681; DB 69; Length 735;
 Best Local Similarity 96.9%; Pred. No. 1.1e-157;
 Matches 712; Conservative 0; Mismatches 21; Indels 2; Gaps 2;

523 TTTGGGATTTTGGTTTCATGTTTGGAGAACCCCTGTCACCAAGAGAAATATTCOA 582
 1 TTTGGGATTTTGGTTTCATGTTTGGAGAACCCCTGTCACCAAGAGAAATATTCOA 60
 583 AGAGAGATGATATTTGTAGATCAGAGATCCTTGGAGAGATATATCAGGAAT 642
 61 AGAGAGATGATATTTGTAGATCAGAGATCCTTGGAGAGATATATCAGGAAT 120
 643 TTTGTGAAGTATGTAACAACCTGTGGCAAGGAGCTCTGGCAAAAGAGATGC 702
 121 TTTGTGAAGTATGTAACAACCTGTGGCAAGGAGCTCTGGCAAAAGAGATGC 180
 703 AATTGCGGCAAGATGCGGACCAACCAAGCTGGGCGCTTCCAATGACCCAG 762
 181 AATTGCGGCAAGATGCGGACCAACCAAGCTGGGCGCTTCCAATGACCCAG 240
 763 GAGGTGCTCTGCGAGCAATGCTTATGTCACCAATGTAATGAAGAGCAAGCTGGA 822
 241 GAGGTGCTCTGCGAGCAATGCTTATGTCACCAATGTAATGAAGAGCAAGCTGGA 300
 823 GTAGAAATAGACCTTGGGCTGAGACGCGATGAGTACCCCTTATGGAAGAGTGAG 882
 301 GTAGAAATAGACCTTGGGCTGAGACGCGATGAGTACCCCTTATGGAAGAGTGAG 360
 883 CCTCAGTGGATGGAGCGCTGAGATTTAGCGTCCCAATCAAGTGTCAACGACCCA 942
 361 CCTCAGTGGATGGAGCGCTGAGATTTAGCGTCCCAATCAAGTGTCAACGACCCA 420
 943 AATTATGAAGAGAGAGATGATTTGACCAAAATGAGAGCTCTGATAGTGAAGTCA 1002
 421 AATTATGAAGAGAGAGATGATTTGACCAAAATGAGAGCTCTGATAGTGAAGTCA 480
 1003 CTGGTGGCTTTGAGATGATATTAATCACTGATGATGTCACCAAGTATATTTCCCGG 1062
 481 CTGGTGGCTTTGAGATGATATTAATCACTGATGATGTCACCAAGTATATTTCCCGG 540

QY	703	AATGTGGCGCAAGAGATGCGGACACCACCGCTGGGCCCTTG6GCCTTCCAATTGACCCAG	762
Db	596	AATTGTGCGCAAGAGATGCGGACACCACCGCTGGGCCCTTG6GCCTTCCAATTGACCCAG	537
QY	763	GAGGTGGCTCCTGCAGGAATGCCCCAATGTCCAACCTAGTGAATGAAAGAACAAGCTGGAA	822
Db	536	GAGGTGGCTCCTGCAGGAATGCCCCAATGTCCAACCTAGTGAATGAAAGAACAAGCTGGAA	477
QY	823	GTACAAATAGACCTGGGGGTGAGAGACGGCATTCGAGTAACCCCTTTATTTGGAAGAGTAG	882
Db	476	GTACAAATAGAGCCTGGGGGTGAGAGACGGCATTCGAGTAACCCCTTTATTTGGAAGAGTAG	417
QY	883	CCTACGCGGATPVGSGGACCTGAGATTTACGGTTCCGAATCAAAGTGTCAAGCACCA	942
Db	416	CCTACGCGGATPVGSGGACCTGAGATTTACGGTTCCGAATCAAAGTGTCAAGCACCA	357
QY	943	ATAATTTGAAGAAGAGAGATGATTTGTACACAAATGTGACAGTCTCATTTAGTTAGTCA	1002
Db	356	ATAATTTGAAGAAGAGAGATGATTTGTACACAAATGTGACAGATCTCATTTAGTTAGTCA	297
QY	1003	CTGCTGGCTTTGAGAGATGATTTACTCATCTTGGATGGTGTCAACAGTTCATTA-TTTTCCG	1061
Db	296	CTGCTGGCTTTGAGATGATTTACTCATCTTGGATGGTGTCAACAGTTCATTAATTTTCCG	237
QY	1062	GGATTAAGTATCCAGGCGCCAGSAGAGGAATTAATGGAAGAAAGGGAAGGCTCCCACATT	1121
Db	236	GGATTAAGTATCCAGGCGCCAGSAGAGGAATTAATGGAAGAAAGGGAAGGCTCCCACATT	177
QY	1122	TGACAAACAAATATCAAGGGCTCTTTGATATATCACTTTTATGTGATTTTCCAAAAGA	1181
Db	176	TGACAAACAAATATCAAGGGCTCTTTGATATATCACTTTTATGTGATTTTCCAAAAGA	117
QY	1182	ACAGTTAACAGAGAGCGAGCGAAGAGATATCAACACGCTACTGAAACAAGGTCACTGCA	1241
Db	116	ACAGTTAACAGAGAGCGAGCGAAGAGATATCAACACGCTACTGAAACAAGGTCACTGCA	57
QY	1242	GAGGTATACAAATGAGCTGCAAGGATTTATTTGAGAGTGAATTAATTTGACCTTTGTTT	1297
Db	56	GAGGTATACAAATGAGCTGCAAGGATTTATTTGAGAGTGAATTAATTTGACCTTTGTTT	1
RESULT	3	AA639658	695 bp mRNA EST 23-OCT-1997
LOCUS		nq83977.s1 NCI-CGAP Co9 Homo sapiens cDNA clone IMAGE:1158972 3'	
DEFINITION		similar to WP:TI1SH9.1 CE01664 DNAJ ;, mRNA sequence.	
ACCESSION		AA639658	
VERSION		AA639658.1 GI:2563437	
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS		Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
TITLE		1 (bases 1 to 695)	
JOURNAL		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
COMMENT		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
		Unpublished (1997)	
		On May 9, 1995 this sequence version replaced gi:802630.	
		Contact: Robert Strausberg, Ph.D.	
		Tel: (301) 496-1550	
		Email: Robert_Strausberg@nih.gov	
		Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,	
		M.D., Ph.D.	
		cDNA Library Preparation: M. Bento Soares, Ph.D.	
		cDNA library Arrayed by: Greg Lennon, Ph.D.	
		DNA sequencing by: Washington University Genome Sequencing Center	
		clone distribution: NCI-CGAP clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LNLB at:	
		www-bio.lnl.gov/bdtp/image/image.html	

FEATURES		Location/Qualifiers
source		1..695
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:1158972"
		/clone.lib="NCI CGAP Co9"
		/tissue.type="colon tumor RER+"
		/lab.host="DH10B"
		/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with 2 modified polylinker; 1st strand cDNA was prepared from RER+ colon tumor, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Ronaldo (Soares4)."
BASE COUNT	153 a	188 c 120 g 234 t
ORIGIN		
Query Match	45.0%;	Score 619.4; DB 36; Length 695;
Best Local Similarity	96.8%;	Pred. No. 1.7e-142;
Matches 673;	Conservative 0;	Mismatches 18; Indels 4; Gaps 4;
619	TTTGAGAAAGTA-TATGCGAAATTTTGTGGAAGTATTAACCAACCTGTGGCAG	677
D6	TT	695
678	GCAGGCTCTGGCAACGGAAAGTCAATTTGTGGCAAGATGCGACACCACTGGG	737
D6	TT	695
635	GCAGGCTCT - GCAAACGGAAAGTCAATTTGTGGCAAGATTTGCGACCACTGGG	577
738	CCCTGGGCGCTTCCAAATATCACAGAGGTGCTTGGCAGATGCGCTATATGCAACT	797
D6	TT	695
576	CCCT - GGCCTTCCAAATATCACAGAGGT - GTCGCGAGATGCGCTATATGCAACT	519
798	AGTAATGAAGAAGCAACGCTGGAAGTAAATATGAGCTGGGGTGAAGACGGATGA	857
D6	TT	695
518	AGTAATGAAGAAGCAACGCTGGAAGTAAATATGAGCTGGGGTGAAGACGGATGA	459
858	GTAACCCCTTATTTGGAAGGTGAGCCTCACGTGATGGGGACCTGGAGATTACGTT	917
D6	TT	695
458	GTAACCCCTTATTTGGAAGGTGAGCCTCACGTGATGGGGACCTGGAGATTACGTT	399
918	CCGAATCAAAAGTTGTCAACACCACTATTTTAAAGAGAGAGATATTTGTACACAA	977
D6	TT	695
398	CCGAATCAAAAGTTGTCAACACCACTATTTTAAAGAGAGAGATATTTGTACACAA	339
978	TGTGACAGCTCATTTAGTTGAGTCACTGGTGGCTTTGAGATGATTTACTCACTTGA	1037
D6	TT	695
338	TGTGACAGCTCATTTAGTTGAGTCACTGGTGGCTTTGAGATGATTTACTCACTTGA	279
1038	TGTGACAGCTCATTTATTTCCCGGGATTAAGATCACAGGCCAGGAGCAANTANTGAA	1097
D6	TT	695
278	TGTGACAGCTCATTTATTTCCCGGGATTAAGATCACAGGCCAGGAGCAANTANTGAA	219
1098	GAAAGGGAGAGGCGCCCACTTTGACAACAACAATATCAAGGGCTTTGATATATCAC	1157
D6	TT	695
218	GAAAGGGAGAGGCGCCCACTTTGACAACAACAATATCAAGGGCTTTGATATATCAC	159
1158	TTTTGATGTGATTTTCCCAAAGAAGTATTAACAGAGACGAGAGAGATATCAACA	1217
D6	TT	695
158	TTTTGATGTGATTTTCCCAAAGAAGTATTAACAGAGACGAGAGAGATATCAACA	99
1218	GCTACTGAACAAGGCTGATGACAGAAAGTATTAACATGACATGCAAGATTTGAGAGTG	1277
D6	TT	695
98	GCTACTGAACAAGGCTGATGACAGAAAGTATTAACATGACATGCAAGATTTGAGAGTG	39
1278	AATAAATTGGACCTTTTAAATTAAGTGAATA 1312	
D6	TT	695
38	AATAAATTGGACCTTTTAAATTAAGTGAATA 4	

LOCUS	AA984302/c	655 bp	MRNA	EST	27-MAY-1998
DEFINITION	am83c01.s1 Strataene schizo brain S11 Homo sapiens cDNA clone IMAGE:1629744 3' similar to WP:115H9.1 CE01664 DNAS ;, mRNA sequence.				
ACCESSION	AA984302				
VERSION	AA984302.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 (bases 1 to 655) Hiller,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S., Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R. WashU-NCI human EST Project				
TITLE	Unpublished (1997)				
JOURNAL	On Sep 12, 1996 this sequence version replaced gi:1404710.				
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu This clone is available royalty-free through LML; contact the IMAGE consortium (info@image.lml.gov) for further information. Possible reversed clone: similarity on wrong strand Seq primer: -40m13 fwd. ET from Amersham High quality sequence scop: 433. Location/Qualifiers 1. 655 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1629744" /clone_lib="Stratagene schizo brain S11" /sex="male" /tissue.type="schizophrenic brain S-11 frontal lobe" /dev_stage="34 years old" /lab_host="SOLR (kanamycin resistant)" /note="Vector: Bluescript SK-; Site_1: EcoRI; library constructed from S-11 frontal lobe, male, 34 years old, 50% caucasian, 50% Aleutian. Schizophrenic suicide. Random primed into EcoRI site of ZAP II Vector. Mass indexed. Custom library. Avg insert length 1.4kb. Material obtained by Johnston N., Torrey, E.F., Yolken R., and the Stanley Neuropathology Consortium - Analysis of RNAs from the Brains of individuals with Psychiatric Diseases (unpublished) Stanley Neuroirology Laboratory, Johns Hopkins School of Medicine, Baltimore MD."				
BASE COUNT	141 a	179 c	126 g	209 t	
ORIGIN					
Query Match	41.6%;	Score 572.8;	DB 40;	Length 655;	
Best Local Similarity	95.1%;	Pred. No. 5.2e-131;			
Matches 622;	Conservative 0;	Mismatches 29;	Indels 3;	Gaps 3;	
631	TATGACGAAATTTTGTGGAAGTACTGTAGAAACAACAACCTGTGGCAGAGCGCTCCTGGC	690			
632		691			
633		692			
634		693			
635	TATCAGCGAAATTTTGTGGCACTATTAAAGAAACAACCTGTGACAAAGCAGCTCCTGGC	694			
636		695			
637		696			
638		697			
639		698			
640		699			
641	AAACGAGATGCAATTTGTCGG-CAAGAGATGC-GGACCAACCCAGCTGGGCGCTGGCGCT	748			
642		749			
643		750			
644		751			
645	AAACGTCAGTGCACATCTGTGCGCCAAAGAGATGCAGAAACCCACCCAGCTGGGCGCTGGCGCT	756			
646		757			
647		758			
648		759			
649		760			
650		761			
651		762			
652		763			
653		764			
654		765			
655		766			
656		767			
657		768			
658		769			
659		770			
660		771			
661		772			
662		773			
663		774			
664		775			
665		776			
666		777			
667		778			
668		779			
669		780			
670		781			
671		782			
672		783			
673		784			
674		785			
675		786			
676		787			

OY		808	GAGCAACGCTGGAAAGTGAATATGAGCCTGGGGTGAGACAGCGCATGTGACTACCCTTT	867
Db		415	GACGACACGCTGGAGTGTGAATATGAGCCTGGGGTGAGACAGCGCATGTGACTACCCTTT	416
OY		868	ATTGGAGAAGGTGAGCCCTCACGTGGATGGGAGGCCCTGGAGATTTTACGGTTCGCATCAA	927
Db		415	ATTGGAGAAGGTGAGCCCTCACGTGGATGGGAGGCCCTGGAGATTTTACGGTTCGCATCAA	356
OY		928	GTTGTCAAGCACCCCAATATTTTTGAAGAAGAGAGAGATGATTTTGTACACAATATGTGACATC	987
Db		355	GTTGTCAAGCACCCCAATATTTGAAAAGGAGAGAGATGATTTTGTACACAATATGTGACATC	296
OY		988	TCCATAGTTGAGTCACTACGTGGCTGGCTTGGCTTGAAGTATTAATCTACTCTGGATGTGCACAA	1047
Db		295	TCATTAGTTGAGTCACTACGTGGCTTGGCTTGAAGTATTAATCTACTCTGGATGTGCACAA	236
OY		1048	GTACATATTTTCCCGGATTAAGTACACACGCGCCAGAGCGAANTNTGGAAGAAAGGGGAA	1107
Db		235	GTACATATTTTCCCGGATTAAGTACACACGCGCCAGAGCGAANTNTGGAAGAAAGGGGAA	176
OY		1108	GGGCTCCCCAACCTTTGACACACACATAATACAGGGCTCTTTGATTAATCACTTTTGATGTG	1167
Db		175	GGGCTCCCCAACCTTTGACACACACATAATACAGGGCTCTTTGATTAATCACTTTTGATGTG	116
OY		1168	GATTTTCCAAAAAGAACACTTTAACAGAGAACGAGAGAGAGATATCAACACACTCTGATA	1227
Db		115	GATTTTCCAAAAAGAACACTTTAACAGAGAACGAGAGAGAGATATCAACACACTCTGATA	56
OY		1228	CAGGGTCAGTGCAGAGAGTATACAAATGAGTGCAGAGGATTTTGAGAGTGAATA	1281
Db		55	CAGGGTCAGTGCAGAGAGTATACAAATGAGTGCAGAGGATTTTGAGAGTGAATA	2
RESULT	5			
LOCUS	AA59885/c			
DEFINITION	aa32e11.s1 Jia Done marrow stroma Homo sapiens cDNA clone IMAGE:1091276 3' similar to WP:rl15H9.1 CE01664 DNAJ ; mRNA sequence.			
ACCESSION	AA59885			
VERSION	AA59885.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Jia,L., Robey,P., Young,M., Schuler,G., Powell,J., Yang,L., Lennon,G., Hillier,L., Allen,M., Bowles,L., Geisel,S., Kucba,T., Myria,T., Matelin,J., Stepien,M., Tan,F., Theising,B., Bowers,Y., Walter,T., Waterston,R., Wilson,R. and Francomano,C.			
TITLE	WashU-MGB/NHGRI EST Project			
JOURNAL	Unpublished (1997)			
COMMENT	On May 9, 1995 this sequence version replaced gi:802280. Contact: Wilson RK / Jia L WashU-MGB/NHGRI EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LINT ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. Possible reversed clone: similarity on wrong strand Insert Length: 1376 Std Error: 0.00 Seq primer: -40ml3 fwd. ET from Amersham High quality sequence stop: 359. Location/Qualifiers			
FEATURES				
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1091276" /clone_id="jia done marrow stroma"			

BASE COUNT	117 a	137 c	82 g	162 t	
ORIGIN	<pre> /sex="mixed" /tissue_type="bone marrow stroma" /dev_stage="mixed" /lab_host="X11-Blue MR"/SOLR" /ncbi_host="vector: bluescript; site:1: EcorI; site:2: XhoI; oligo-dt priming. Directionally cloned. Size-selected for average insert size >0.5 kb. Library supplied by Dr. Libin Jia (NHRRI)." </pre>				
Query Match	35.2%	Score 484.8;	DB 36;	Length 498;	
Best Local Similarity	98.2%;	Pred. No. 2.5e-109;			
Matches 489;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0;	
OY	799	GGGATGGAAGAACGACGCTGGACATAGAAATAGACCCGCGGCTGAGAGAACGGCATGGAG	858		
Db	498	GGGAATGGAAGACCGGACCGCTGGAAATAGAAATAGACCTGGGTGGAACCGGCATGGAG	439		
OY	859	TACCCCTTATTTGGAGAGAGTGAGGCTCACGTCGATGGAGGACCTTGAGATTTAAGGTTTC	918		
Db	438	TACCCCTTATTTGGAGAGAGTGAGGCTCACGTCGATGGAGGACCTTGAGATTTAAGGTTTC	379		
OY	919	CGAATCAAAAGTTGTCAAGCACCCCAATATTTGAAAGAGAGAGATGATTGTACACAAAT	978		
Db	378	CGAATCAAAAGTTGTCAAGCACCCCAATATTTGAAAGAGAGAGATGATTGTACACAAAT	319		
OY	979	GTGACAGTCTCATTTAGTTGATGACACTGTGTGGCTTTGAGATGGATTTACTACTTGGAT	1038		
Db	318	GTGACAGTCTCATTTAGTTGATGACACTGTGTGGCTTTGAGATGGATTTACTACTTGGAT	259		
OY	1039	GGTCAAGAGGTACATATTTCCGGGGTAATAGATCACAGGCCAGAGAGCGAANTANTGGAAG	1098		
Db	258	GGTCAAGAGGTACATATTTCCGGGGTAATAGATCACAGGCCAGAGAGCGAANTANTGGAAG	199		
OY	1099	AAAGGGGAAGGGCTCCCACTTTGACACAACAATATCAAGGGCTCTTTGATTAATCACT	1158		
Db	198	AAAGGGGAAGGGCTCCCACTTTGACACAACAATATCAAGGGCTCTTTGATTAATCACT	139		
OY	1159	TTTGTATGTGATTTTCCAAAAGAACAGTTAACAGAGAGAACGAGACGATCAACACAG	1218		
Db	138	TTTGTATGTGATTTTCCAAAAGAACAGTTAACAGAGAGAACGAGACGATCAACACAG	79		
OY	1219	CTACGAAACAAGGGTCACTGTCGAGAAGTATATCAATGACCTCAAGAGATTTGAGAGTGA	1278		
Db	78	CTACGAAACAAGGGTCACTGTCGAGAAGTATATCAATGACCTCAAGAGATTTGAGAGTGA	19		
OY	1279	ATMAAATTGGACTTTGTT	1296		
Db	18	ATMAAATTGGACTTTGTT	1		
RESULT	6	AA581995/c			
LOCUS	AA581995	488 bp	EST	05-SEP-1997	
DEFINITION	nm36d05.s1 NCI_CGAP GC5 Homo sapiens cDNA clone IMAGE:1085961	3'			
ACCESSION	AA581995				
VERSION	AA581995.1	GI:2360673			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 488)				
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
	Tumor Gene Index				
	Unpublished (1997)				
	Contact: Robert Strausberg, Ph.D.				
	Tel.: (301) 496-1550				

Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/INLNL at:
 www-bio.lnl.gov/bdrrp/image/image.html

Seq primer: -40m13 fwd. ER from Amersham
 High quality sequence stop: 344.

FEATURES

Source

Location/Qualifiers

1. 488

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1085961"

/clone_1lb="NCI-CGAP-GC5"

/tissue_type="germ cell tumor"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site: 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. Mixed
germ cell tumors. 5' adaptor sequence: 5' GAATTCGGCAGCAG
3' 3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3'

Average insert size: 0.7 kb.

Average insert size: 0.7 kb.

BASE COUNT 115 a 135 c 80 g 158 t
 ORIGIN

Query Match 35.0%; Score 481.2; DB 35; Length 488;
 Best Local Similarity 99.0%; Pred. No. 1.9e-108;

Matches 483; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 811 CGAAGCTGAGTAGAATAAGACCTGGGGTGAGACAGCGCATGAGTACCCCTTATT 870

DB 488 CGAAGCTGAGTAGAATAAGACCTGGGGTGAGACAGCGCATGAGTACCCCTTATT 429

QY 871 GGAAGAGTACCTCAGTGGATGGGAGCCTGGAGATTACGGTCCGAATCAAGTT 930

DB 428 GGAAGAGTACCTCAGTGGATGGGAGCCTGGAGATTACGGTCCGAATCAAGTT 369

QY 931 GTCAGAGACCCATATTGTAAGAGAGAGATGATTGACAAATGCAAGTCTCA 990

DB 368 GTCAGAGACCCATATTGTAAGAGAGAGATGATTGACAAATGCAAGTCTCA 309

QY 991 TTAGTTGAGTCACTGGTGGCTTTGAGATGATATTACTGCTGATGCTCAAGGTA 1050

DB 308 TTAGTTGAGTCACTGGTGGCTTTGAGATGATATTACTGCTGATGCTCAAGGTA 249

QY 1051 CATATTCCCGGATTAAGATCAGCAGCCGAGGAGCAANTANTGAAAGAGGAAAGG 1110

DB 248 CATATTCCCGGATTAAGATCAGCAGCCGAGGAGCAANTANTGAAAGAGGAAAGG 189

QY 1111 CTCGCCAAGCTTGGACAACAATATCAAGGCTCTTTGATATCACTTTTGTGTGAT 1170

DB 188 CTCGCCAAGCTTGGACAACAATATCAAGGCTCTTTGATATCACTTTTGTGTGAT 129

QY 1171 TTTCGAAAAGACGTTAAACAGAGAGAGAGAGAGATCAACAGCTACTGAAACAA 1230

DB 128 TTTCGAAAAGACGTTAAACAGAGAGAGAGAGAGATCAACAGCTACTGAAACAA 69

QY 1231 GGGTCGTGAGAGATTAATCAATGATGCAAGGATATTGAGAGTAAATTTGAC 1290

DB 68 GGGTCGTGAGAGATTAATCAATGATGCAAGGATATTGAGAGTAAATTTGAC 9

QY 1291 TTTGTTTA 1298

DB 8 TTTGTTTA 1

RESULT 7
 N93316/c N93316 482 bp mRNA EST 20-AUG-1996
 LOCUS

DEFINITION

z69e03.s1 Soares_fetal_lung_NbHL19W Homo sapiens CDNA clone
 IMAGE:308860 3' similar to WP:TI5H9.1 CE01664 DNAJ ; mRNA
 sequence.

ACCESSION N93316 GI:1265625

VERSION EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT: Willson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through INLNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Insert Length: 1849 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 470.

Location/Qualifiers

1. 482

/organism="Homo sapiens"

/db_xref="Gene:1252273"

/clone="IMAGE:308860"

/clone_1lb="Soares_fetal_lung_NbHL19W"

/dev_stage="19 weeks"

/lab_host="DHI0B (ampicillin resistant)"

/note="Organ: lung; Vector: pT73D (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer.

15'-TGTTCAATCTGAAGTGGAGGCGCGCAATTTTCTTTTCTTTT-3',

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot - 5. Library constructed by Bento

Soares and M. Fatima Bonaldo. This library was constructed

from the same fetus as the fetal heart library, Soares

fetal heart NbHL19W."

BASE COUNT 115 a 134 c 78 g 155 t

ORIGIN

Query Match 34.7%; Score 476.8; DB 25; Length 482;
 Best Local Similarity 99.2%; Pred. No. 2.3e-107;

Matches 478; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 817 CTGGAAGTAGAATAGAGCTGGGGTGAGAGAGCGCATGAGATACCCCTTATTGGAGAA 876

DB 482 CTGGAAGTAGAATAGAGCTGGGGTGAGAGAGCGCATGAGATACCCCTTATTGGAGAA 423

QY 877 GGTGAGCTCACGTGGATGGGAGGCTGAGATTACGGTCCGAATCAAGTTGTCAAG 936

DB 422 GGTGAGCTCACGTGGATGGGAGGCTGAGATTACGGTCCGAATCAAGTTGTCAAG 363

QY 937 CACCCATATTGGAAGAGAGAGATGATTGTACACAAATGTGACAGCTCATTTAGTT 996

DB 362 CACCCATATTGGAAGAGAGAGATGATTGTACACAAATGTGACAGCTCATTTAGTT 303

```

OY      997 GAGTCACTGTTGGCTTGGAGTATATTAATCTGATGGTGCACAGGTACATATT 1056
         |||||||
Db      302 GAGTCACTGTTGGCTTGGAGTATATTAATCTGATGGTGCACAGGTACATATT 243

OY      1057 TCCCGGGATTAAGATCACCAGGCCAGAGCGAATFANTGAAGAAAGGGAGGCTCCCC 1116
         |||||||
Db      242 TCCCGGGATTAAGATCACCAGGCCAGAGCGAATFANTGAAGAAAGGGAGGCTCCCC 183

OY      1117 AACCTTGACAACAATATCAAGGGCTCTTGTGATTAATCACTTTTGGATGGATTTTCCA 1176
         |||||||
Db      182 AACCTTGACAACAATATCAAGGGCTCTTGTGATTAATCACTTTTGGATGGATTTTCCA 123

OY      1177 AAAGAACGTTAAACAGAGAGAGCGAGATATCAACAGCTCTGAACCAAGGTC 1236
         |||||||
Db      122 AAAGAACGTTAAACAGAGAGAGCGAGATATCAACAGCTCTGAACCAAGGTC 63

OY      1237 GTGCAAGAGTATACAAATGACATCGAAGATATTGAGAGTGAATAAATTTGACTTTGTT 1296
         |||||||
Db      62 GTGCAAGAGTATACAAATGACATCGAAGATATTGAGAGTGAATAAATTTGACTTTGTT 3

OY      1297 TA 1298
         ||
Db      2 TA 1

RESULT 8
A1819362/c 791 bp mRNA EST 24-AUG-1999
LOCUS      w61b11.x1 Soares NSF_F8_9W_OT_PA_P.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2369565.3 similar to WP:T15H9.1 CE01664 HYPOTHETICAL. 39.9 KD
            PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR ;, mRNA sequence.
ACCESSION  A1819362
VERSION     A1819362.1 GI:5438441
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 791)
            NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            On Jun 5, 1998 this sequence version replaced gi:3189482.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.lnl.gov) for further information.
            Seq primer: -400p from GlDco
            High quality sequence stop: 469.
            Location/Qualifiers
                1..791
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2369565"
                /clone_id="Soares_NSF_F8_9W_OT_PA_P.S1"
                /lab_host="DH10B"
                /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
                a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                Equal amounts of plasmid DNA from five normalized
                libraries were mixed, and ss circles were used as tracer in
                a subtractive hybridization reaction. The driver was
                PCR-amplified cDNAs from pools of 5,000 clones made from
                the same 5 libraries. The pools consisted of the following
                libraries and clones: Soares NBHSF pool 1:
                309384-310919, 323208-325895 Soares NB2HP pool 1:
                145032-147335, 147720-148103, 148872-149255, 15002 -
                150407, 151176-152327 Soares NB2HF8-9W pool 1:
                758280-760583, 772104-774407 Soares NBHPA pool 1:
                304776-306311, 320136-322823, 326280-326663 Soares NBHOT
                304776-306311, 320136-322823, 326280-326663 Soares NBHOT

```

```

BASE COUNT      234 a      186 c      114 g      254 t      3 others
ORIGIN
Query Match      33.1%; Score 455.6; DB 61; Length 791;
Best Local Similarity 95.0%; Pred. No. 4.5e-102;
Matches 478; Conservative 0; Mismatches 24; Indels 1; Gaps 1;
OY      874 GAAGGTGAGCCCTCAGCTGATGGAGCCCTGAGATATTACGGTTCGAATCAAGTTGTC 933
         |||||||
Db      791 GAAGGTGAGCCCTCAGCTGATGGAGAGCTGAGATTTACGGTTCGAATCAAGTTGTC 732

OY      934 AAGCACCCATATTTTGAAGAGAGAGATGATTTTGTACACAATGTGACATCTCATTA 993
         |||||||
Db      731 AAGCACNCATTA-TTGAAGAGAGAGATGATTTTGTACACAATGTGACATCTCATTA 673

OY      994 GTTGAGTCACTGGTGGCTTTGAGATGGATATTACCACTTGATGTCACAAGGTACT 1053
         |||||||
Db      672 GTTGAGTCACTGGTGGCTTTGAGATGGATATTACCACTTGATGTCACAAGGTACT 613

OY      1054 ATTTCCCGGATAGATCACAGCCAGAGCGAANTANTGGAAGAAAGGGAGGCTC 1113
         |||||||
Db      612 ATTTCCCGGATAGATCACAGCCAGAGCGAANTANTGGAAGAAAGGGAGGCTC 553

OY      1114 CCCAACTTTGACACACATATATCAAGGGCTCTTGTGATTAATCACTTTGATGTGATTTT 1173
         |||||||
Db      552 CCCAACTTTGACACACATATATCAAGGGCTCTTGTGATTAATCACTTTGATGTGATTTT 493

OY      1174 CCMAAAGAACATTTAAACAGAGAGAGAGAGAGATATCAACACTCTACTGAAACAAGG 1233
         |||||||
Db      492 CCMAAAGAACATTTAAACAGAGAGAGAGAGAGATATCAACACTCTACTGAAACAAGG 433

OY      1234 TCAGTGCAGAAAGTATACATGAGCTCAAGAGATATTGAGATGAATAAATTTGACTTT 1293
         |||||||
Db      432 TCAGTGCAGAAAGTATACATGAGCTCAAGAGATATTGAGATGAATAAATTTGACTTT 373

OY      1294 GTTTAAATTAAGTGAATTAAGCATATTTATATCTGCAGAGTTTGTGTGTGTTTTT 1353
         |||||||
Db      372 GTTTAAATTAAGTGAATTAAGCATATTTATATCTGCAGAGTTTGTGTGTGTTTTT 313

OY      1354 GTTTTATTTTCAATATGCAACT 1376
         |||||||
Db      312 GTTTTATTTTCAATATGCAAGT 290

RESULT 9
AAB37534/c 767 bp mRNA EST 18-MAR-1998
LOCUS      oes1e05.s1 NCI-CCAP_Pr25 Homo sapiens cDNA clone IMAGE:1410176
DEFINITION similar to WP:T15H9.1 CE01664 DNMF; contains Alu repetitive
            element;., mRNA sequence.
ACCESSION  AAB37534
VERSION     AAB37534.1 GI:2912733
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 767)
            NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            On Jan 17, 1998 this sequence version replaced gi:1900092.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright,
            Ph.D.
            cDNA Library Preparation: Stratagene, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.

```

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnlnl.gov/bdip/image/image.html

Insert Length: 1639 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 487.
Location (contig) 487

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TMAGE:1410176"
/clone_1b="NCI_CGAP_P-25"
/tissue_type="epithelium (cell line)"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: prostate; Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: oligo dT. Normal prostate epithelial cell line (HIV immortalized). 5' adaptor sequence: 5' GAATTCGACACGAC 3' adaptor sequence: 5' CTCGACATTTTTTTTTTTTTTTTTTT 3'
Average insert size: 1.1 kb.

```

BASE COUNT	227 a	177 c	114 g	249 t
ORIGIN				

Query Match	32.48;	Score	445.4;	DB	39;	Length	767;
Best Local Similarity	96.18;	Pred. No.	1.5e-99;				
Matches	466;	Conservative	0;	Mismatches	18;	Indels	1;
						Gaps	1

OY	892	GATGGGAGACCTGGAGATTTACGGTTCGGAATCAAAAGTGTCAAGACCAACCAATATTGAA	951
Db	767	GATGGGGACCTGGAGATTTACGTTCCGATCAAGGTGTCAAGCCCAACCAATATTGAAA	708
OY	952	AGGAGAGAGATGATTTGTACACAAATGTGACAGTCTCATTTAGTGAAGTCTGGTTGGC	1011
Db	707	AGGAGAGAGATGATTTGTACACAAATGTGACAAATCTCATTTAGTGAAGTCTGGTTGGC	648
OY	1012	TTTGAGATGATATTACTCACTTGGATGGTCACAAAGTACATTTTCCCGGATTAAGATC	1077
Db	647	TTTGAGATGATATTACTCACTTGGATGGTCACAAAGTACATTTTCCCGGATTAAGATC	589
OY	1072	ACCAGGCCAGGAGCGCAANTANTGGAMAAAGGGGAAGGGCTCCCAACTTTGACAACAAC	11331
Db	588	ACCAGGCCAGGAGCGAAGCTATGAMAAAGGGGAAGGGCTCCCAACTTTGACAACAAC	529
OY	1132	AATATCAAGGGCTCTTTGATATAACACTTTGATGTGGATTTTCCAAAGAACAGTTAA	11991
Db	528	AATATCAAGGGCTCTTTGATATAACACTTTGATGTGGATTTTCCAAAGAACAGTTAA	469
OY	1192	GAGGAGCGAGAGAGGTATCAACACGCTACTGAACAAGGGTCAAGTGCAGAGAGTATAC	1251
Db	468	GAGGAGCGAGAGAGAGTATCAACACGCTACTGAACAAGGGTCAAGTGCAGAGAGTATAC	409
OY	1252	AATGAGCTGCAAGGATTTTGAAGTCAATAAATTTGGACTTTGTTTAAATTAAGTGAATA	13111
Db	408	AATGAGCTGCAAGGATTTTGAAGTCAATAAATTTGGACTTTGTTTAAATTAAGTGAATA	349
OY	1312	AGCGATATTATTATTCGCAAGGTTTTTTTGTGTGTTTTGTTTTATTTCATTAATG	13711
Db	348	AGCGATATTATTATTCGCAAGGTTTTTTTGTGTGTTTTGTTTTATTTCATTAATG	289
OY	1372	CAAGT	1376
Db	288	CAAGT	284

RESULT	10
AM323622	
LOCUS	556 bp mRNA
DEFINITION	EST
	26-JAN-2000
	U074606.y1 NCI C639 Mami Mus musculus cDNA clone IMAGE:2648290 5'
	similar to WP.T15H9.1 CE01664 HYPOTHETICAL 39.9 KD PROTEIN T15H9.1
	IN CHROMOSOME II PRECURSOR ;, mRNA sequence.

ACCESSION	AM323622
VERSION	AM323622.1
KEYWORDS	GI:6757647
SOURCE	EST.
	house mouse.

ORGANISM	REFERENCE	AUTHORS	TITLE
<i>Mus musculus</i>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (base 1 to 556)	NCI-CCGAP	http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP)

MG1:1028742
Seq primer: -40RP from Gibco
High quality sequence stop: 304.
Location/Qualifiers

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="2648290"
/clone_1lb="NCL-CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; site.1: SalI;
site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

```

BASE COUNT	156 a	110 c	161 g	129 t
ORIGIN				
Query Match	32.4%	Score 445.2;	DB 80;	Length 556;
Best Local Similarity	87.7%	Pred. NO. 1.5e-99;		
Matches 466; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0;

[illegible]

713	AAGAGATGCGGACACACCACAGCTGGGCGCCCTGGGGCGCTTCCAAATGACACCCAGGAGGTGGCT	772
Db	1 AAGGATGAAACCCACACAGCTGGGACCGGGACGGCTTCCAAATGACACCCAGGAGGTGGCT	60
Qy	773 GCGACGATGCCCTTAATGTCAAACTAGTAAATGAAGAAACGAAACGCTGGAGTAGAATAG	832
Db	61 GTGACGAGTCCCTTAATGTCAAACTAGTAAATGAAGAAACGAAACCTAGATGGAATAG	120
Qy	833 AGCCTGGGCTGAGAGACGGCATGAGTACCCCTTTATTTGGAGAAAGTGAGCTCAGCTG	892
Db	121 AGCCTGGGCTGAGAGTGGCATGAGTACCCCTTTATTTGGAGAAAGTGAGCCCATGTGG	180
Qy	893 ATGGGGACCCCTGGAGATTACGGTTCCGAATCAAAAGTTGTCAACACCCATATTTTAAA	952
Db	181 ATGGGGACCCCGGAGACTTACCGTTCCGAATCAAAAGTTGTCAACACCCGATTTTAGA	240
Qy	953 GGAGAGAGATGATTTTGTATCACAAATGTACAGCTCATTTAGTTGAGTACATCGTTGGCT	1012
Db	241 GGAGAGGAGATGACGTGTACACAAATGTACAGCTCATCGTTGAGGCTGTGGTGGCT	300
Qy	1013 TTGAGATGATTAATTAATCACTTGATGGATGGTCCACAGGTACATATTTTCCGGGATTAATCA	1072
Db	301 TTGAGATGAGCATATTAATCACTTGATGGATGGTCCACAGGTACATATTTTCCGGGACAAATCA	360
Qy	1073 CCAGGCCAGAGGGAANTTNTGGAGAAAGGGGAGGCGCTCCCAACTTTGACACAAACA	1132
Db	361 CCAGGCCAGAGGCCAAGCTGTGGAGAAAGGGGAGGCTGCCCAACTTTTATTAACATA	420
Qy	1133 ATATCAAGGCGCTTTTGATTAATCACTTTTGATGTGATTTTCCAAAGAACAGTTAACAG	1192
Db	421 ACATCAAGGCGCTTTTGATTAATCACTTTTGATGTGATTTTCCAAAGAACAGCTGACAG	480
Qy	1193 AGGAGCGAGAGAAAGGTATCAAAACAGCTACTGAAACAAGGCTCAGTGAGA	1243
Db	481 AGGAGCGAAAGAAAGGTATCAAGACCTTTCTCAACAAAGGCGCGGTGAGA	531
RESULT 12		
LOCUS	A1057091	750 bp mRNA EST 01-OCT-1998
DEFINITION	o223e10.x1 Soares.total.fetus_Nb2HF8.9w Homo sapiens cDNA clone	
	IMAGE:1676202 3' similar to WP:115H9.1 CE01664 DNASJ ;, mRNA	
	sequence.	
ACCESSION	A1057091	
VERSION	A1057091.1	GI:3330967
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 750)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP)	
JOURNAL	Unpublished (1997)	
COMMENT	On May 18, 1995 this sequence version replaced gi:811070. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.lln.gov) for further information. Insert Length: 862 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 445. Location/Qualifiers 1..750 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1676202" /clone_lib="Soares_total.fetus_Nb2HF8.9w" /dev_stage="8-9 weeks" /lab_host="DH10B"	
FEATURES		
source		


```

RESULT 14
LOCUS   A1800795/c
DEFINITION
A1800795      726 bp      mRNA      EST      19-DEC-1999
w913c05.x1 Soares.NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2364968 3' similar to WP:115H9.1 CE01664 HYPOTHETICAL 35.9 KD
PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR ;, mRNA sequence.
ACCESSION A1800795
VERSION   A1800795.1 GI:5366267
KEYWORDS EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 726)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL   Unpublished (1997)
COMMENT    On Feb 22, 1999 this sequence version replaced gi:4283012.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1350
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1027 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 445.
Location/Qualifiers
1. 726
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2364968"
/lab_host="Soares.NSF_F8_9W_OT_PA_P_S1"
/note="Organ: pooled: Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker: Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOR
pool 1: 723720-726407, 739080-740999 Subtraction by Benito
Soares and M. Fatima Bernaldo."
BASE COUNT 215 a 164 c 104 g 241 t 2 others
ORIGIN
Query Match 30.3%; Score 417; DB 60; Length 726;
Best Local Similarity 96.8%; Pred. No. 1,5e-92;
Matches 423; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 940 CCAATATTGAAAGGAGAGATGATTTGTACCAAAATGACAGCTCATATTAGTAGAG 999
Db 726 CCCAAATATTGAAAGGAGAGATGATTTGTACCAAAATGACAAATCTTATGATGAG 667
QY 1000 TCACTGGTGGCTTTGAGATGATATTACTACTTGGATGGTGCACAGGTACATATTCC 1059
Db 666 TCACTGGTGGCTTTGAGATGATATTACTACTTGGATGGTGCACAGGTACATATTCC 607
QY 1060 CGGAGTAAGTACACAGGCCAGAGCGAANTANTGAGAAAGGGAAGGGCTCCCAAC 1119
Db 606 GGGGATTAAGTACACAGGCCAGAGCGAGCTATGAGAAAGGGAAGGGCTCCCAAC 547
QY 1120 TTGACACAAACATATCAAGGGCTCTTGTATATCACTTTTGTATGATGATTTCCAAA 1179

```

```

Db 546 TTGACAAACAAATATCAAGGGCTCTTGTATATCACTTTTGTATGATGATTTCCAAA 487
QY 1180 GAACAGTTAAACAGAGGAAGGAGAGGTATTCAAACAGCTACTCGAAACAGGGTCACTG 1239
Db 486 GAACAGTTAAACAGAGGAAGGAGAGGTATTCAAACAGCTACTCGAAACAGGGTCACTG 427
QY 1240 CAGAAGGTATCAATGAGCTCAAGGATATGAGAGTGAATAAATGAGTCTTTGTTAA 1299
Db 426 CAGAAGGTATCAATGAGCTCAAGGATATGAGAGTGAATAAATGAGTCTTTGTTAA 367
QY 1300 AATAAGTAATTAAGCGATATTTATCTGCAAGGTTTTTTGTGTGTTTTGTTTTT 1359
Db 366 AATAAGTAATTAAGCGATATTTATCTGCAAGGTTTTTTGTGTGTTTTGTTTTT 307
QY 1360 ATTTTCATATGCAAGT 1376
Db 306 ATTTTCATATGCAAGT 290

RESULT 15
LOCUS   AM122551/c
DEFINITION
AM122551      480 bp      mRNA      EST      22-OCT-1999
U1-M-BH2.2-sox-d-09-0-U1.s1 NIH_BMAP_M_S3.2 Mus musculus cDNA clone
U1-M-BH2.2-sox-d-09-0-U1 3', mRNA sequence.
ACCESSION AM122551
VERSION   AM122551.1 GI:6098075
KEYWORDS EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 480)
AUTHORS   Bernaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
discovey
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
COMMENT    On Feb 22, 1999 this sequence version replaced gi:4283549.
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mbest@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares lab clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA-No.
Location/Qualifiers
1. 480
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U1-M-BH2.2-sox-d-09-0-U1"
/clone_1lb="NIH_BMAP_M_S3.2"
/dev_stage="27-32 days"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker: Site.1: Not I; Site.2: Eco RI; The
NIH_BMAP_M_S3.2 library is a subtracted library of a
series, ultimately derived from a mixture of individually
tagged normalized libraries from ten regions of the mouse
brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus) after a series of
subtractions to reduce the representation of cDNAs from
which ESTs had already been generated. The following
serially subtracted libraries were generated in this

```

process: NIH_BMAP_M.S3.2, NIH_BMAP_M.S2, NIH_BMAP_M.S1.
The subtracted library (NIH_BMAP_M.S3.2) was constructed
as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2
clones from which 3' ESTs had been derived was used as a
driver in a hybridization with the NIH_BMAP_M.S2 library
in the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (liferotechnologies) to generate the
NIH_BMAP_M.S3.2 library. This procedure has been
previously described (Bonaldi, Lennon and Soares, Genome
Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M.S3.2
TAG_TISSUE=basal-ganglia
TAG_SEQ=TCGAC

BASE COUNT 110 a 128 c 112 g 130 t
ORIGIN

Query Match 30.2%; Score 415.4; DB 69; Length 480;

Best Local Similarity 92.4%; Pred.No. 3.3e-92;

Matches 437; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
OY 198 GGCCTCCGACAGAACCTGAGCACCCTTTGCTGCTGCTATACCTCATCGGGCGGTGAT 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 GGCCTCCGACAGAACCTGAGCACCCTTTGCTGCTGCTATACCTCATCGGGCGGTGAT 421

OY 258 TGGCGAGCAGATTTCTATAGATCTTGGGGGCTGCTCGAAGTGCCTCTATAAGATAT 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 CGCGGGGCGGATTTCTATAAGATCTTGGGGGCTGCTCGAAGTGCCTCCATTAAGGACAT 361

OY 318 TAAAGGCTATAGGAACCTAGCCTTCAGCTTCAATCCCGACCGGAACCTGATGTC 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TAAAGGCTATAGGAACCTAGCCTTCAGCTTCAATCCCGACCGGAACCTGATGTC 301

OY 378 ACAAGCCAGAGAAATTCAGAGATCTGGGTGCTGCTTATGAGGTTCTGTACAGTAGTA 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CCAAGCCAGAGAAATTCAGAGATCTGGGTGCTGCTTATGAGGTTCTGTACAGTAGTA 241

OY 438 GAAACGGAACAGTACGATCTATGTGAAGAGATTAAGATGTCATCAGAGCTC 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 AAAACGGAACAGTACGATCTATGTGAAGAGATTAAGATGTCATCAGAGCTC 181

OY 498 CCATGAGACATTTTTCACACTTCTTGGGATTTTGGTTTCATGTTTGAGAGAACCC 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 TCATGGGACATTTTTCACACTTCTTGGGATTTTGGTTTCATGTTTGAGAGAACCC 121

OY 558 TCGTCAGCAGACAGAATATTCAGAGAGAGATGATATTATTGTAGATCTAGAGTCA 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 TCGTCAGCAGACAGAATATTCAGAGAGAGATGATATTATTGTAGATCTAGAGTCA 61

OY 618 TTGGAGAAGTATATGACAGAAATTTTGGGAAGTAGTTAGAACAACACTG 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 TCTGGAAGAAGTATATGACAGAAATTTTGGGAAGTAGTTAGAACAACACTG 8
```

Search completed: June 30, 2000, 12:12:13
JOB time: 8064 sec

127 NIPRGSIDIVDLEVTLEEYAGNFEEVVRNKPVARQAPGKRKCNCRQEMRTTQLGPGRFQ 186

Db 126 GTPKADVTIDLFVLEEVNGHEVEIKRRAVYKOTSGTRQCNCRHMRTEOMQGRFQ 185
 QY 187 MTOEVCDECPNKLVTNEETLEVEIEBPGVDGMEYPTIGGEPHVOGEPPDLNPRIVV 246
 Db 186 MFQKVCDECPNKLVTNEETLEVEIEBPGVDGMEYPTIGGEPHVOGEPPDLNPRIVV 245
 QY 247 KHPFERGDDLYTNVTVSLVESLVGFEMDITLHDKRHVHSRDKTPRGAXXMKKGEG 306
 Db 246 KHPFERGDDLYTNVTVSLVESLVGFEMDITLHDKRHVHSRDKTPRGAXXMKKGEG 305
 QY 307 PNFNNNNKSLITFPVDFPKREOTTEBARGITKOLKQSGVQ-KVINGL 355
 Db 306 PSLDNNKGMVYTFVEFPKTELSDOKAOITLILQONTVKRKAAYNGL 355

RESULT 2

DNJ2_ALLPO STANDARD; PRT; 418 AA.

AC P42824;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNJ2
 GN DNJ2
 OS Allium portum (leek).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eukaryotes; Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;
 OC Alliaceae; Allium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Besoulet J.J., Testet E., Cassagne C.;
 RT "Cloning of a new isoform of a DnaJ protein from Allium portum
 RL epidermal cells."
 RT Plant Physiol. Biochem. 32:723-727(1994).
 CC -1- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY
 CC IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS A DNJ2-LIKE DOMAIN AND THE FOUR CXKCGXG
 CC MOTIFS FOUND IN DNJ2 PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: X77632; CAA54720.1; -
 DR HSSP: P25685; 1HDJ.
 DR PRINTS: PR00625; DNJ2PROTEIN.
 DR PROSITE: PS00636; DNJ2_1; 1.
 DR PROSITE: PS00637; DNJ2_CXKCGXG; 1.
 DR PROSITE: PS00076; DNJ2_2; 1.
 DR PFAM: PF00226; DnaJ_1.
 DR PFAM: PF00684; DnaJ_CXKCGXG; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 KW Chapterone; Prenylation; Lipoprotein; Multigene family.
 KM DOMAIN 11 76 DNJ2-LIKE.
 FT DOMAIN 83 88 POLY-GLY.
 FT DOMAIN 81 108 GLY-RICH.
 FT REPEAT 148 155 CXKCGXG MOTIF.
 FT REPEAT 164 171 CXKCGXG MOTIF.
 FT REPEAT 191 198 CXKCGXG MOTIF.
 FT REPEAT 207 214 CXKCGXG MOTIF.
 FT LIPID 415 415 FARNESYL (BY SIMILARITY).
 SQ SEQUENCE 418 AA; 46584 MW; DCE2AADI192329E6 CRC64;

Query Match 29.2%; Score 549.5; DB 1; Length 418;
 Best Local Similarity 35.0%; Pred. No. 6.4e-38;
 Matches 126; Conservative 75; Mismatches 118; Indels 41; Gaps 10;

QY 26 FYKILGTPRSASINDIKATKRLALQHPDNPDPDQAEKFDLLGAAYVLSDESKRQ 85
 Db 14 YVEVLGSKNATPDLKAVRKAKIKNPKRGD---PKFKEIGAYEVLNDPEKREI 69
 QY 86 YDTGEGGLDGHOSHSG-----DIFSHFGDFGFMGTPROODRIIPGSDIYDLE 140
 Db 70 YDQYGEGLGKMGSGGSGVNDPFDIPOSFTFGGGGFGGGGSSR--GRORRGGEDVHPK 127
 QY 141 TLEEVYAGNEVEYRNPV-ARQAPGK-----KRCNCR-----QEMRTTOLGPGRFMT 188
 Db 128 SLEDLYNGTSGKLSLSHNVLCTKCKGSKSGASMCASQSGSGMKVYSIRQLGPMIQM 187
 QY 189 Q-----EYV-----CDECPNKLVTNEETLEVEIEBPGVDGMEYPTIGGEPHYD 233
 Db 188 QHPONCKGTGEMISDKRCPDCKGKGVQKVLVHVKRGMONGKITTFPEADPADP 247
 QY 234 GERDLEFRKIVKHPFERGDDLYTNVTVSLVESLVGFEMDITLHDKRHVHSRDK 291
 Db 248 TVTGDYIVLQKQKHEHFRKKGDDLEFHSLSLEALCGFVTLHDKRHVHSRDK 307
 QY 292 ITPRGAXXMKKGGLPFDNPNKSLITFPVDFPKREOTTEBARGITKOLKQSGVQ 351
 Db 308 VIKPDQFKGINDGEMPTQRPFRKGLYHFSVDFP-DSLTPDCKALBSVLPFRNASRL 366

RESULT 3

DNH2_ARATH STANDARD; PRT; 419 AA.

AC P42825; Q43293;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE DNJ2
 GN DNJ2
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-CV. COLUMBIA;
 RA Zhou R., Hayman G.T., Kroczyńska B., Mierzyński J.A.;
 RL Submitted (Xxx-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 81-192 FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Cooke R., Laudie M., Raynal M., Delsey M.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY
 CC IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS A DNJ2-LIKE DOMAIN AND THE FOUR CXKCGXG
 CC MOTIFS FOUND IN DNJ2 PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: L36113; AAB86799.1; -
 DR EMBL: F20032; CAA23386.1; -
 DR HSSP: P25685; 1HDJ.
 DR PRINTS: PR00625; DNJ2PROTEIN.
 DR PROSITE: PS00636; DNJ2_1; 1.
 DR PROSITE: PS00637; DNJ2_CXKCGXG; 1.
 DR PROSITE: PS00076; DNJ2_2; 1.
 DR PFAM: PF00226; DnaJ_1.
 DR PFAM: PF00684; DnaJ_CXKCGXG; 1.
 DR PFAM: PF01556; DnaJ_C; 1.

RESULT	4			
CC	DNJH_CUCSA	STANDARD;	PRT;	413 AA.
AC	004960;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	DNMJ PROTEIN HOMOLOG (DNMJ-1).			
GN	DNMJ1.			
OS	Cucumis sativus (Cucumber).			
CC	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
CC	euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
CC	cory eudicots; Rosidae; eurosids I; Cucurbitales; Cucurbitaceae;			
CC	Cucumans.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=SEED;			
RX	MEDLINE; 93343632.			
RA	Preisig-Mueller R., Kindl H.;			
RT	"Plant dnaJ homologue: molecular cloning, bacterial expression, and			
RT	expression analysis in tissues of cucumber seedlings.";			
RL	Arch. Biochem. Biophys. 305:30-37(1993).			
CC	-I- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY			
CC	IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS.			
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN SEEDLINGS IN ALL TISSUES, BUT			
CC	EXCESSIVELY HIGH LEVELS IN HYPOCOTYLEDONS AND ROOTS.			
CC	-I- INDUCTION: WEAK, BY HEAT SHOCK			
CC	-I- SIMILARITY: CONTAINS A DNMJ-LIKE DOMAIN AND THE FOUR CXXCGXG			
CC	MOTIFS FOUND IN DNMJ PROTEINS.			
CC	-----			

[illegible]

RESULT	5			
ID	DNH1_ATTRNU	STANDARD:	PRT:	417 AA.
AC	P43644:			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	DNM1 PROTEIN HOMOLOG ANJ1.			
OS	Atirplex nummulata.			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Caryophyllales; Caryophyllales; Chenopodiaceae;			
OC	Atirplex.			
RP	[1]			
RP	SEQUENCE FROM N.A.			


```

RX MEDLINE: 93222693.
RA Zhu J.R., Shi J., Bressan R.A., Hasegawa P.M.:
RT "Expression of an Atirplex nummularia gene encoding a protein
RT homologous to the bacterial molecular chaperone DnaJ."
RL Plant Cell 5:341-349(1993).
CC
CC -1- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY
CC IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -1- INDUCTION: BY HEAT SHOCK, AND IN RESPONSE TO NaCl STRESS.
CC -1- SIMILARITY: CONTAINS A DnaJ-LIKE DOMAIN AND THE FOUR CXKCGXG
CC MOTIFS FOUND IN DnaJ PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L09124; -; NOT_ANNOTATED_CDS.
DR HSSP: P25685; 1HDJ.
DR PRINTS: PR00625; DnaJPROTEIN.
DR PROSITE: PS00637; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXKCGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
DR PFAM: PF01556; DnaJ_C; 1.
KM Chaperone; Repeat; Prenylation; Lipoprotein.
FT DOMAIN 11 76 DnaJ-LIKE.
FT DOMAIN 81 106 GLY-RICH.
FT DOMAIN 81 106 POLY-GLY.
FT REPEAT 148 155 CXKCGXG MOTIF.
FT REPEAT 164 171 CXKCGXG MOTIF.
FT REPEAT 191 198 CXKCGXG MOTIF.
FT REPEAT 207 214 CXKCGXG MOTIF (APPROXIMATIVE).
FT LIPID 414 414 FARNESYL (BY SIMILARITY).
SQ SEQUENCE 417 AA; 46566 MW; BA597AEBA3A22667 CRC64;

Query Match 28.0%; Score 526.5; DB 1; Length 417;
Best Local Similarity 34.9%; Pred. No. 4.9e-36;
Matches 123; Conservative 70; Mismatches 118; Indels 41; Gaps 11;

OY 26 FYIILGVPNSASIKDKIKKAYRLALDLPDNDPOAOKFODLGAAYEVLSDSEKRO 85
DB 14 YVEILGVPKDSPELDLKKAYKKAIRKHPDKGD---PEKFEKLAHAYEVLSDPEKREI 69
OY 86 YDYYGEGGKLDGQSSHG-----DIFSHFFGDFGFMFGGTPROQDNRIPGSDIIVDLEY 140
DB 70 YDYGEDALKEGSGGGGGMHDPFDIFQSFEG--GSPFEGVSSGRGRORGEVDVHPLKY 127
OY 141 TLEEVYAGNEVEY-VRNKPYARQAPGK-----RKC-N-CR--O-MRTTQIGPGRFQMT 188
DB 128 SLEDLTGTTKKLSLRNVICSKCTGKSGSKSGSCGCGMAYSTRHLPSPNTIOM 187
OY 189 Q-----EVY-----CDECPNVKLYNEERTLEVELEPGVRDMEYPTIGEGPHVD 233
DB 188 QHPCNCKGTGETINDKRCPOCKGEKVQOEKVLVEYVEKGMQHQKITFPGDEAD 247
OY 234 GEPDGLRFITKVKHPFERRGDDLTNTYVSLVESLVFEMDITLHDHAKYHISD--K 291
DB 248 YTVGLDVLVLOQKEHKPKFRKGEDELYEHTLSLTLEALCFRVLVTHLDGRLLISNLGE 307
OY 292 TTRPGAXXKKKKEGLPNFDNNIKGSLITFDVDPKPEQGLTEAREGIRQL 343
DB 308 VVKPDQFKAIEDGMPYGRPFMKGMKYHFTVEFP-DSLNDVQKSLFAIL 358

RESULT 6
DNU1_ALLPO STANDARD: PRT; 397 AA.

```

```

AC 003363;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE DnaJ PROTEIN HOMOLOG 1 (DnaJ-1) (FRAGMENT).
GN DnaJ.
OS Allium porrum (leek).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;
OC Alliaceae; Allium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93265942.
RA Bessoule J.T.;
RT "Occurrence and sequence of a DnaJ protein in plant (Allium porrum)
RT epidermal cells."
RL FEBS Lett. 323:51-54(1993).
CC -1- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY
CC IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -1- SIMILARITY: CONTAINS A DnaJ-LIKE DOMAIN AND THE FOUR CXKCGXG
CC MOTIFS FOUND IN DnaJ PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69436; CAA49211.1; -
DR PIR: S31387; S31387.
DR PIR: S33312; S33312.
DR HSSP: P25685; 1HDJ.
DR PROSITE: PS00637; DnaJ_1; 1.
DR PROSITE: PS00637; DnaJ_CXKCGXG; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXKCGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
KM Chaperone; Repeat; Prenylation; Lipoprotein; Multigene family.
FT NON_TER 1 1
FT DOMAIN 1 52 DnaJ-LIKE.
FT DOMAIN 62 66 POLY-GLY.
FT DOMAIN 62 66 GLY-RICH.
FT REPEAT 127 134 CXKCGXG MOTIF.
FT REPEAT 143 150 CXKCGXG MOTIF.
FT REPEAT 170 177 CXKCGXG MOTIF.
FT REPEAT 186 193 CXKCGXG MOTIF.
FT LIPID 394 394 FARNESYL (BY SIMILARITY).
SQ SEQUENCE 397 AA; 44169 MW; 33BA6B981586EFAB CRC64;

Query Match 27.6%; Score 519.5; DB 1; Length 397;
Best Local Similarity 35.6%; Pred. No. 1.7e-35;
Matches 124; Conservative 66; Mismatches 117; Indels 41; Gaps 9;

OY 34 NSASIKDKIKKAYRLALDLPDNDPOAOKFODLGAAYEVLSDSEKROYTYGEEG 93
DB 1 KNASPDLLKKAIRKKAIRKHPDKGD---PEKFEKLAQAVDVLSDPEKREIYQYGEDA 56
OY 94 LKDGQSSHG-----DIFSHFFGDFGFMFGGTPROQDNRIPGSDIIVDLEYLVEYVAG 148
DB 57 LKEGSGGGGCHDHPDIFDIFQSFEGGGGSGSSR--GHRORGEVDVHPLKVSLEELNG 114
OY 149 NFEVYRNKPY-----ARQAPGKRCNCR--O-MRTTQIGPGRFQMTQ----- 190
DB 115 TSKKLSLRNVICSKCNGKSGSKSGASCOSGSKMVSIRGLGPMIDQMOPNDCK 174
OY 190 ---EVY-----CDECPNVKLYNEERTLEVELEPGVRDMEYPTIGEGPHVDGGLRF 241
DB 175 GTGEMINDKRCPLCKEKKVQOEKVLVEYVEKGMQGORITFPGDEADPDTVTGDIVF 234

```



```

OY 242 RIKVVKHPIFERGGDLTYNTVTVSLVESLVGFEMDITHLDGKHVHSRD--KITRGAXX 299
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 235 VLDQKQHPFKFORKGDDLFYKHTLSLALCGFQFVLTHLDGRQLLTKSNGEYVVKPDQFK 294
OY 300 WKKGEGLPNFDNNNNKSLITFDVDFPKQEQLEAREGKILLKOGS 347
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 295 AINDEBMPYQRPFRMGKXLIQFLVDFP-DSLTPDCKVLESVLPNSA 341

RESULT 7
DNJ2_HUMAN STANDARD; PRT; 397 AA.
AC P31689;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNJ2 PROTEIN HOMOLOG 2 (HSJ-2).
GN HSJ2 OR DNJ2 OR HDJ2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP MEDLINE: 93326630.
RA Oh S., Iwahori A., Kato S.;
RT "Human cDNA encoding Dnaj protein homologue.";
RL Biochim. Biophys. Acta 1174:114-116(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93326629.
RA Chellalal A., Davis A., Mohanakumar T.;
RT "Cloning of a unique human homologue of the Escherichia coli Dnaj
RT heat shock protein.";
RL Biochim. Biophys. Acta 1174:111-113(1993).
CC -1 SIMILARITY: CONTAINS A DNJ-LIKE DOMAIN AND THE FOUR CXCCGXG
CC MOTIFS FOUND IN DNJ PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D13388; BAA02656.1; -
DR EMBL: L08069; AAC37517.1; -
DR PIR: S34631; S34631.
DR HSSP: P25685; 1HDJ.
DR MIM: 602837; -
DR PRINTS: PR00625; DNJPROTEIN.
DR PROSITE: PS00636; DNJ_1; 1.
DR PROSITE: PS00637; DNJ_CXCCGXG; 1.
DR PROSITE: PS50076; DNJ_2; 1.
DR PFAM: PF00226; Dnaj; 1.
DR PFAM: PF00684; Dnaj_CXCCGXG; 1.
DR PFAM: PF01556; Dnaj_C; 1.
KW Chapterone; Repeat.
FT DOMAIN 4 70 DNJ-LIKE.
FT REPEAT 134 141 GLY-RICH.
FT REPEAT 150 157 CXCCGXG MOTIF.
FT REPEAT 177 184 CXCCGXG MOTIF.
FT REPEAT 193 200 CXCCGXG MOTIF.
FT REPEAT 274 274 Q -> H (IN REF. 1).
FT CONFLICT 274 274
SQ SEQUENCE 397 AA; 44868 MW; A899C0F6BB32780 CRC64;

Query Match 27.4%; Score 515.5; DB 1; Length 397;
Best Local Similarity 34.5%; Pred. No. 3,7e-35;
Matches 133; Conservative 66; Mismatches 117; Indels 69; Gaps 13;

```

```

OY 26 FYKILGVRASASTIKDIKAYRKALALDHPDNDPQAOEKFODLGAAYVLSDSERKQ 85
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 7 YVDLVGKPNATQDEELKAYRKALALKHPDKNPE--GEKFOISQAYVLSDAKKRL 63
OY 86 YDYEGLKLDGHDQ---SSGDIFFSHFDFGFMFGTFRQDORNPRESDIIVDLEVY 141
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 64 YDKGGEALKEGGAGGFGSGPMDIFDMFEG-----GGGMOKER--RKNVYHQLSVT 114
OY 142 LEEVYAGNFEVYVNRKPV---ARQAPKR--KC--NCR--QEMFTTQLGPGRFQMTQ 189
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115 LEDLYNGATRKIALQKNVICDKCEGRCGKGAVECCPCNCRGTQMRHIGCMVQIOI 174
OY 190 EYV-----CECPNVKLVNEERLVEIEEGVADGMETPIIGGEPIVGD 234
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 175 SVMECEGHERISPKDRCSCKMGKIVREKILLEVHIDKMGDQKITFHGGEDEPGL 234
OY 235 EPGLDRIRIKVYVPIFERGGDLTYNTVTVSLVESLVGFEMDITHLDGKHVHSR--DKI 292
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 235 EPGLDIIVLQKDHAVYTRGEGDLFMCMQIQLVEALCGFKPISTLDNRTIVTSHGQI 294
OY 293 TRGAXXWKKGEGLPNFDNNNNKSLITFDVDFPK-----EQLTEAREGIR 340
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 295 VKHGDIKCVLNEGMPYRYREKGRITIEFKVNPENGFISPKLSLEKLPERKE-VE 353
OY 341 QLKQGSVQRY-----YNG 354
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 354 ETDEMDELVELVDFPDQERRRRYNG 378

RESULT 8
DNJ2_MOUSE STANDARD; PRT; 397 AA.
AC P54102;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNJ2 PROTEIN HOMOLOG 2 (HSJ-2).
GN HSJ2 OR DNJ2 OR RDJ1.
OS Mus musculus (Mouse), and Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES-MOUSE;
RX MEDLINE: 99017984.
RA Royaux I., Minner F., Goffinet A.M., de Rouvroit C.L.;
RT "A Dnaj-like gene, Hsj2, maps to mouse chromosome 5, at approximately
RT 24 cM from the centromere.";
RL Genomics 53:415-415(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-RAT; STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RA Leng C.H., Wang C.;
RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: CONTAINS A DNJ-LIKE DOMAIN AND THE FOUR CXCCGXG
CC MOTIFS FOUND IN DNJ PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF055664; AAC78597.1; -
DR EMBL: U53922; AAA98855.1; -
DR HSSP: P25685; 1HDJ.
DR MGD: MGI:1270129; HSJ2.
DR PRINTS: PR00625; DNJPROTEIN.
DR PROSITE: PS00636; DNJ_1; 1.
DR PROSITE: PS00637; DNJ_CXCCGXG; 1.
DR PROSITE: PS50076; DNJ_2; 1.

```

DR PFAM: PF00226; DnaJ_1.
 DR PFAM: PF00684; DnaJ_CXXCXXG; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 KW Chaperone; Repeat.
 FT DOMAIN 4 70 DNAJ-LIKE.
 FT DOMAIN 75 96 GLY-RICH.
 FT REPEAT 134 141 CXXCXXG MOTIF.
 FT REPEAT 150 157 CXXCXXG MOTIF.
 FT REPEAT 177 184 CXXCXXG MOTIF.
 FT REPEAT 193 200 CXXCXXG MOTIF.
 SQ SEQUENCE 397 AA; 44868 MW; 1783CE3D5C4CD558 CRC64;

Query Match 27.4%; Score 515.5; DB 1; Length 397;
 Best Local Similarity 34.5%; Pred. No. 3.7e-35;
 Matches 133; Conservative 66; Mismatches 117; Indels 69; Gaps 13;

QY 26 FYKILGVPRASIKDIKAYRKALQLHPDNPDPQAQEKFDLGAAYEVLSDSEKRRQ 85
 DB 7 YVDLVGVPRNATQELKAYRKALKYHPDKNPNE--GEKFKQISQAYEVLADSKREL 63
 QY 86 YDTGEEGLKDGHQ-----SSHGDIFFSHFGDFGFMFGSTPRQODRNIPRGSDDITVDEVT 141
 DB 64 YDKGGEQAIKSGAGGSGSPMDFDFG-----GGGRMQRR--RGKNVHQLSVT 114
 QY 142 LEEVYAGNEVEVVRNKPV---AAQAPGR--KC--NCR--OEMRTTOLGPRFQMTQ 189
 DB 115 LEDVYNGATRKALQKNVICDKCEGRGKGAECPCNCGTQMIHQGPWMQIQ 174
 QY 190 EVV-----CDCCPVKVLNERTLEVELEPVGRGMETPRIGSEEPVHG 234
 DB 175 SVCEECQCHGRISPKRSCNCRKLTREKILVHIDGMKQGRITFRGEDDQEPGL 234
 QY 235 EPGDLERIKVKKHIFERRGDDLYTNVTVSLVESLVGFEMDITHLGCHVHTSR--DKI 292
 DB 235 EPGDITVLDQKHAVFTRGEDLFGMCDQLVLCGFGKPISTLNNRTVITSHGQI 294
 QY 293 TRPAXXMKKGEGLPNDNNIKSLITFDVDFK-----EQLTEARBGIR 340
 DB 295 VKHGDICVLNEGPIRYRREYKRLIEFKVNPENGLSPDKLSLEKILPERKE--VE 353
 QY 341 QLLKQSVQKV-----YNG 354
 DB 354 ETDEMDEVLEVPDPNERRRHNG 378

RESULT 9
 MASS_YEAST
 ID MASS_YEAST STANDARD; PRT; 409 AA.
 AC P25491;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MITOCHONDRIAL PROTEIN IMPORT PROTEIN MASS (PROTEIN YD1).
 GN MASS OR YD1 OR YNL064C OR N2418 OR YNL2418C.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92107179.
 RA Atencio D.P., Yaffe M.P.;
 RT "MASS, a yeast homolog of DnaJ involved in mitochondrial protein
 RT import.";
 RL Mol. Cell. Biol. 12:283-291(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 91332099.
 RA Caplan A.J., Douglas M.G.;
 RT "Characterization of YD1: a yeast homologue of the bacterial dnaJ
 RT protein.";
 RL U. Cell Biol. 114:609-621(1991).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1676;
 RX MEDLINE; 96021608.
 RA Berge P., Doignon F., Crouzet M.;
 RT "The sequence of a 44 420 bp fragment located on the left arm of
 RT chromosome XIV from Saccharomyces cerevisiae.";
 RL Yeast 11:967-974(1995).
 RN [4]
 RP ERRATUM.
 RX MEDLINE; 97060022.
 RA Berge P., Doignon F., Crouzet M.;
 RL Yeast 12:297-297(1996).
 RN [5]
 RP ISOPRENOLD.
 RX MEDLINE; 92406811.
 RA Caplan A.J., Tsai J., Casey P.J., Douglas M.G.;
 RT "Purification of YD1p is required for function at elevated growth
 RT temperatures in Saccharomyces cerevisiae.";
 RL J. Biol. Chem. 267:18890-18895(1992).
 CC -1- FUNCTION: PROBABLY INVOLVED IN MITOCHONDRIAL PROTEIN IMPORT. IS
 CC ALSO REQUIRED FOR EFFICIENT TRANSLLOCATION OF PRE-PRO-ALPHA-FACTOR.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED IN A PERINUCLEAR RING AS WELL
 CC AS IN THE CYTOPLASM (ACCORDING TO REF.2).
 CC -1- INDUCTION: MASS IS A HEAT SHOCK GENE WHOSE EXPRESSION INCREASES
 CC MODERATELY AT ELEVATED TEMPERATURES.
 CC -1- SIMILARITY: CONTAINS A DnaJ-LIKE DOMAIN AND THE FOUR CXXCXXG
 CC MOTIFS FOUND IN DnaJ PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S74758; AAB20771.1; -;
 DR EMBL; X56560; CAA39910.1; -;
 DR EMBL; U12141; AAA99647.1; -;
 DR EMBL; 271340; CAA95937.1; -;
 DR PIR; A39659; A39659.
 DR PIR; S26703; S26703.
 DR HSSP; P25685; 1HDJ.
 DR SGD; L0002503; YD1.
 DR PRINTS; PR00625; DnaJPROTEIN.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS00637; DnaJ_CXXCXXG; 1.
 DR PROSITE; PS00765; DnaJ_2; 1.
 DR PFAM; PF00226; DnaJ_1.
 DR PFAM; PF00684; DnaJ_CXXCXXG; 1.
 DR PFAM; PF01556; DnaJ_C; 1.
 KW Chaperone; Heat shock; Transport; Protein transport; Mitochondrion;
 KW Repeat; Prenylation; Lipoprotein.
 FT DOMAIN 4 72 DnaJ-LIKE.
 FT DOMAIN 75 103 GLY-RICH.
 FT REPEAT 143 150 CXXCXXG MOTIF.
 FT REPEAT 159 166 CXXCXXG MOTIF.
 FT REPEAT 185 192 CXXCXXG MOTIF.
 FT REPEAT 201 208 CXXCXXG MOTIF.
 FT LIPID 406 406 FARNESYL.
 SQ SEQUENCE 409 AA; 44670 MW; E4539F3618DD9CF2 CRC64;

Query Match 27.1%; Score 510; DB 1; Length 409;
 Best Local Similarity 36.6%; Pred. No. 1.1e-34;
 Matches 130; Conservative 56; Mismatches 119; Indels 50; Gaps 14;

QY 26 FYKILGVPRASIKDIKAYRKALQLHPDNPDPQAQEKFDLGAAYEVLSDSEKRRQ 85
 DB 7 YVDLVGVPRNATQELKAYRKALKYHPDKNPSE-EAAEKFKASAYEILSDPEKRD 65
 QY 86 YDTGEEGLKDGHSN-----DIFSHFGDFGFMFGSTPRQODRNIPRGSDDIT 135
 ||:||||| | : | ||||| | |

```

Db 66 YDQEGEDL-SGAGAGGPGGFGFDLPSPFFGAGG-----AQRPGRQKDK 117
Qy 136 VDLEVTLEEVYAGNFVEVNRKPV-ARQAPK-----RKC-NCRDE--MRTTQACP-183
Db 118 HEISASLEELKGTATLAKNQLCKECEGRGKGAIVKCTSCNGGIGFVTRQMPM 177
Qy 183 -GFRQ-----TQEVV-----CDECPNVKLVNEERTLEVEIPGVADGMEYPIGGE 229
Db 178 IQRQTECDVCHGTGDIIDPKDRCKSCNGKVKENERKLEHVHVGMDGGRIVFKGAD 237
Qy 230 PHVDGEEDLFRIRKVKHPFERRGDLYTNVTVSLVESLVGFEMDITHLDGH--KVHI 287
Db 238 QAPVDIVGDVFIYSESRPKSFGRDGLVYEAFIDLTALAGEFPALEHSGMLAKGI 297
Qy 288 SRDKITPRGAXXMKKGEGLPNEDNNNKGSLITFPDVPKQQLTEARREGIKQL 342
Db 298 VPGEVIAPGRMKVIEKGMP-IPRYGYGNLIFKTIKFPENHFTSE--EMLKRL 349

RESULT 10
DNAB_BROUV STANDARD; PRT; 375 AA.
AC 005980;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNAB PROTEIN.
GN DNAB.
OS Brucella ovis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25840 / 63/290T;
RX MEDLINE; 93094135.
RA Cellier M.F.M., Teyssier J., Nicolas M., Liautard J.P., Marti J.,
RA Sri Midada J.;
RT "Cloning and characterization of the Brucella ovis heat shock protein
RT dnaK functionally expressed in Escherichia coli.";
RL J. Bacteriol. 174:8036-8042(1992).
CC - FUNCTION: INTERACTS WITH DNAB, TO DISASSEMBLE A PROTEIN COMPLEX
CC AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY
CC WITH GREP, THE ATPASE ACTIVITY OF DNAB (BY SIMILARITY).
CC - INDUCTION: BY HEAT SHOCK.
CC - SIMILARITY: TO OTHER PROKARYOTIC DNAB, AND TO EUKARYOTIC
CC DNAB-LIKE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M95799; AAC36133.1; -
DR EMBL; M94063; AAA93047.1; -
DR PIR; B47042; B47042.
DR HSSP; P08622; IXL.
DR PRINTS; PR00625; DNABPROTEIN.
DR PROSITE; PS00636; DNAB_1; 1.
DR PROSITE; PS50076; DNAB_2; 1.
DR PROSITE; PS00637; DNAB_CXXCXGXG; 1.
DR PFAM; PF00684; DnaJ_CXXCXGXG; 1.
DR PFAM; PF01556; DnaJ_C; 1.
KM Chapterone; DNA replication; Heat shock; Repeat.
FT DOMAIN 2 71 DNAB-LIKE.
FT DOMAIN 77 106 GLY-RICH.
FT REPEAT 146 153 CXXCXGXG MOTIF.
FT REPEAT 163 170 CXXCXGXG MOTIF.
FT REPEAT 185 192 CXXCXGXG MOTIF.
FT REPEAT 199 206 CXXCXGXG MOTIF.

```

```

SQ SEQUENCE 375 AA; 41021 MW; 61CB5F99FA2DC73D CRC64;
Query Match 24.9%; Score 469; DB 1; Length 375;
Best Local Similarity 32.8%; Pred. No. 2,36-31;
Matches 118; Conservative 59; Mismatches 129; Indels 54; Gaps 12;

Qy 25 DFYKILGPPRSASIKDKATRKALQLHPDRNPDPQAQKFPDGLAAIYVLSSEK 84
Db 4 DYELALGVTRADKTKLAAFRKLAHQYHPDRNPDPQAERKFRKIGAYETLKDPQKRA 63
Qy 85 QYDFYGEGLDKHQ-----SSHGDFSHFSDFFPMGFGPFRDNRNIPRSD 133
Db 64 AYDRGHAALFENGGMGSGFGAGFAGAGGFADIFEDIFE--AMGG--RRRRNGAERKAD 118
Qy 134 IIVDLEVTLEEVYAGNFVEV-----VANKP-----VARQAPGRKRCNCRQEM 175
Db 119 LRYNMEVTLLEEVYAGKTKQIRVPSITCDESGSARKGSPPTCTMGSGGRCRAQGF 178
Qy 176 ----RTTOLGPRRQMTQEVVCDCECPNVKLVNEERTLEVEI-----EPGYRDNMEYPI 225
Db 179 FSVERTCPGCRGQIILD-PCKKCHGQGRYDRGRSISVNIIPAVSRREPRIR-----LA 231
Qy 226 GEGEPHVDPGEF-GDLRFRIKVKHPFERRGDLYTNVTVSLVESLVGFEMDITHLGSK 284
Db 232 GEGEGLRGGRAGDLYILSVKPHFEPORDADLYCKVPSMTTTRALGQEVSTLDSTQ 291
Qy 285 VHSRDKITPRGAXXMKKGEGLPNEDNNNKGSLITFPDVPKQQLTEARREGIKQLK 344
Db 292 TRVKVPECTQNGKQRLKGMKMPGL-RQSVTGDLYIQDIEFP-QNLSKQRELEEFEX 349

RESULT 11
DNAB_ECOLI STANDARD; PRT; 375 AA.
AC P08622;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNAB PROTEIN.
GN DNAB OR GRP.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.
RX MEDLINE; 8611849.
RA Ohki M., Tamura F., Nishimura S., Uchida H.;
RT "Nucleotide sequence of the Escherichia coli dnaJ gene and
RT purification of the gene product.";
RL J. Biol. Chem. 261:1778-1781(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 8611850.
RA Bardwell J.C.A., Tilly K., Craig E., King J., Zyllicz M.,
RA Georgopoulos C.;
RT "The nucleotide sequence of the Escherichia coli K12 dnaJ gene. A
RT gene that encodes a heat shock protein.";
RL J. Biol. Chem. 261:1782-1785(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 92334977.
RA Yusa T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 97426617.

```

```

Db      61 KRAAYDQYGHAAFEQGGGGGGFGGAGDFSDIPEDYVGDIIFGGGRGQ--RAARGADLRY 118
QY      137 DLEVYLEEYVAGNEVEYVANKPVARQA-----PEKRCNCROEMRTTOLGPFROM 187
      119 NMELTLEEARVGVYKEI--RIPTLEEDCDVCHGSAKPGTQPTQ-----PTCHSGQGYQM 171
QY      188 IQ-----EYVCDECPNKLVENEETLEVELEPEYVRGMEYEPFEGS 228
Db      172 RQGFPAVQOTCPHCQGRITLTKDPCNCKCHGGRERKSTLSVKIPACVDTDGRIIRLAGEG 231
QY      229 EPHVDGEF-GDLRRRIKVVYKHPFERRGDILYVTVSLVESLVEGFEMDITLIDGKHVHI 287
      232 PAGEHGAPAGSLYOVQYKQPIFEREGNNLYCEVPINFMAALGEIEVPLDG-RVKL 280
Db      288 SROKTRFGCAXXMKGGSLPEFDNNNNIKGSLITFDVDF-----KEQGLEARE 337
      291 KVPGETQVQGLFPRMGKGVKSV-RGAQGDGLCNVYETVETGINERKQLOLLOLEI 345

```

	RESULT	12			
ID	DNAJ_SALTY		STANDARD;	PRT:	378 AA.
AC	Q60004;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	DNAJ PROTEIN.				
GN	DNAJ.				
OS	Salmonella typhimurium.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-LT2:				
RA	Stephen R.J., Hinton J.C.D.;				
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.				
CC	-I- FUNCTION: INTERACTS WITH DNAJ, TO DISASSEMBLE A PROTEIN COMPLEX				
CC	AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY				
CC	WITH GRPE, THE ATPASE ACTIVITY OF DNAJ.				
CC	-I- SUBUNIT: HOMODIMER (BY SIMILARITY).				
CC	-I- INDUCTION: BY HEAT SHOCK UNDER THE CONTROL OF THE HTPR REGULATORY				
CC	PROTEIN (BY SIMILARITY).				
CC	-I- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC				
CC	DNAJ-LIKE PROTEINS.				
CC	-----				
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@lsb-sib.ch).				
CC	-----				
EMBL:	U58360; AAB02911.1; -.				
DR	HSSP; P08622; IxBL.				
DR	STYLENE; SG7???				
DR	PRINTS; PR00625; DNAJPROTEIN.				
DR	PROSITE; PS00636; DNAJ_1; 1.				
DR	PROSITE; PS50076; DNAJ_2; 1.				
DR	PROSITE; PS00637; DNAJ_CXKXGKG; 1.				
DR	PFAM; PF00226; DnaJ; 1.				
DR	PFAM; PF00684; DnaJ_CXKXGKG; 1.				
DR	PFAM; PF01556; DnaJ_C; 1.				
DR	Chaperone; DNA replication; Heat shock; Repeat.				
FT	INIT MET	0			
FT	DOMAIN	0			
FT	DOMAIN	76	71		
FT	REPEAT	146	112		
FT	REPEAT	163	153	Gly-RICH.	
FT	REPEAT	185	170	CXKXGKG MOTIF.	
FT	REPEAT	192	192	CXKXGKG MOTIF.	
FT	REPEAT	199	206	CXKXGKG MOTIF.	
SEQUENCE	378 AA: 41181 MW: 176ECC64696F30A2 CRC64;				

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF007813; AAC35417.1; -
 DR EMBL: U72647; AAB17396.1; -
 DR HSSP: P08622; 1XBL.
 DR PROSITE: PS00636; DnaJ_1; 1.
 DR PROSITE: PS00076; DnaJ_2; 1.
 DR PROSITE: PS00637; DnaJ_CXXCXXG; FALSE_NEG.
 DR PFAM: PF00226; DnaJ; 1.
 KW Chapterone; DNA replication; Heat shock; Repeat.
 FT DOMAIN 3 72
 FT REPEAT 150 157 CXXCXXG MOTIF.
 FT REPEAT 167 174 CXXCXXG MOTIF.
 FT REPEAT 189 196 CXXCXXG MOTIF.
 FT REPEAT 203 210 CXXCXXG MOTIF.
 SO SEQUENCE 369 AA; 39939 MW; 408B5B11457DCDF CRC64;

Query Match 23.4%; Score 441.5; DB 1; Length 369;
 Best Local Similarity 31.0%; Pred. No. 4.1e-29;
 Matches 117; Conservative 64; Mismatches 135; Indels 61; Gaps 12;

QY 24 RDYKYLIGVRSASINDIKAYRKALQLPDRNPDPQAEKFDGLGAAYEVLSDSEKR 83
 DB 4 RSTYDLIGVSKSANDERIKSAVYRKALKYHPDKNGKSESEKTEATEYELIRDPKRR 63
 QY 84 KOYDTYGEELK-----DGHOSHGDIFSHFGDF-----GFMEGTPRODDRNIPRG 131
 DB 64 QAVDQEGKAGVSGAGGFGGATDSDIRGDLFGDFGSGRSSGGGRRSPQ---RG 120
 QY 132 SDIIVDLVLEVTLEEVYAGN--FVEVVR-----NKPYARQAPGRKRC-NCRQEMRTTQLG 181
 DB 121 SDRYVNLVSLIEDNALGREGYKIEIPRLSECVDSGASGSSPAPCPDGG-----G 172
 QY 182 PGRFOMTQEVY-----CDGCPNVKLVNERTELEVELEPGVRDGMEX 222
 DB 173 SEQIRRTQGFSSVATTCPTCRGKGTIISNFCRSGGGLQEKRRRTINIKIRPVEVTSRL 232
 QY 223 PRTGEGEPHYDEP-GDLRFRIKVVKHPIFERRGDLYTNVTVSLVESLVGFMDITHL 281
 DB 233 KVSGBEAGNGSPHGDLYVVTIKKHELEFRGNDLIVRKISLGAALIGALIEVPTID 292
 QY 282 GKHVHISRDKITRPGAXXMKKGGGLPNFDNNNIKSLITFDVDFPREQLTEPAREGIKO 341
 DB 293 GKAKKKIEGTESGGVFLRKGHGMPLGAYG-KGDQHVIVKIEIPK-KITRRQRELIEE 350
 QY 342 LKQ-----GSYQVY 352
 DB 351 FARESENIPIGSKITF 367

RESULT 15
 DnaJ_METH
 ID DnaJ_METH STANDARD: PRT: 376 AA.
 AC 027352;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE DnaJ PROTEIN.
 GN DnaJ OR MTH1291.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE: 98037514.

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldridge T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadofora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,
 RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Nollan J., Reeve J.N.,
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.",
 RU J. Bacteriol. 179:7135-7153(1997).
 CC -1- FUNCTION: STIMULATES, JOINTLY WITH GRPE, THE APPASE ACTIVITY OF
 CC DNAK (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC DnaJ, AND TO EUKARYOTIC
 CC DnaJ-LIKE PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF000894; AAB85773.1; -
 DR HSSP: P08622; 1XBL.
 DR PRINTS: PR00625; DnaJPROTEIN.
 DR PROSITE: PS00636; DnaJ_1; 1.
 DR PROSITE: PS00076; DnaJ_2; 1.
 DR PROSITE: PS00637; DnaJ_CXXCXXG; FALSE_NEG.
 DR PFAM: PF00226; DnaJ; 1.
 DR PFAM: PF00684; DnaJ_CXXCXXG; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 KW Chapterone; DNA replication; Heat shock; Repeat.
 FT DOMAIN 2 71
 FT REPEAT 148 155 CXXCXXG MOTIF.
 FT REPEAT 165 172 CXXCXXG MOTIF.
 FT REPEAT 191 198 CXXCXXG MOTIF.
 FT REPEAT 205 212 CXXCXXG MOTIF.
 SO SEQUENCE 376 AA; 41584 MW; 3497C7DA61DC5342 CRC64;

Query Match 23.4%; Score 441; DB 1; Length 376;
 Best Local Similarity 34.3%; Pred. No. 4.6e-29;
 Matches 126; Conservative 55; Mismatches 128; Indels 58; Gaps 14;

QY 21 IAGRDYKILGVRSASINDIKAYRKALQLPDRNPDPQAEKFDGLGAAYEVLSDS 80
 DB 1 MARQDYELIGVDRGADKKIKAYRRLARKYHPDVS-DDPAAEKFEISRAYVLSDD 59
 QY 81 ERKQYDTYGEELKNG-----HOSHGDIFSHFGDF-----FMFGTPRODDRN 128
 DB 60 EKRARDRGHGM-DGSEQEDIFNNINFEIDTFSGLGFDIGLFDMGFGGRRHGPQ--- 117
 QY 129 PRGSDIIVDLVLEVTLEEVYAGNVEV---VRNKPV---ARQAPGRKRCNCR-----Q 173
 DB 117 -RGADISTYLDISLEDAYNGLETDIVRPTKCPVCHGSRAEGGTGRCQCGSGGQVR 175
 QY 174 EMRTYDLPGRQMTQ-----EYVCDECPNVKLVNERTELEVELEPGVRDG 219
 DB 176 QVANTTLG---QMNNTTCDPCDCEGIVVEKPCSNCKNGGVVRRKTSITHVVPAGVETG 231
 QY 220 MEYPTIGEPEPV-DQEPDGLFRIRKVVKHPIFERRGDLYTNVTVSLVESLVGFMDIT 278
 DB 232 SRLRVPGEEMGLRGCEPDLYVVIKVKHSTFRREGANLITEKPTISFYQALGLTVEVP 291
 QY 279 HLDGKHVHISRDKITRPGAXXMKKGGGLPNFDNNNIKSLITFDVDFPREQLTEPAREG 338
 DB 292 TLD-RPVKLRIPAGTQSGTFVVKHGMHLMKMGY-GMLYKXVAVVPRKLSPPQ---- 346
 QY 339 IKQLAKQ 345
 DB 346 -KELLRE 351

Sat Jul 1 21:34:08 2000

us-09-501-714-1.rsp

Page 11

Search completed: June 30, 2000, 11:56:22
Job time: 7137 sec

THIS PAGE BLANK (USPTO)

Db 307 VKPQGFKAIDEGMPYGRPFMKGLKTYHFVSVEPPDSLNEQCKALGGV--LPPRTSVQ 363

RESULT 5
T07371
dnaJ protein homolog - potato
C:Species: Solanum tuberosum (potato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C:Accession: T07371
R:Legewie, G.; Braun, H.P.
Plant Physiol. 117, 1127, 1998
A:title: A cDNA from potato with homology to dnaJ is identical to a hitherto unidentified
A:Reference number: Z15992
A:Accession: T07371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-419 <LEG>
A:Cross-references: EMBL:X94301; NID:g1125690; PIDN:CAA6395.1; PID:e214633
A:Experimental source: tissue type leaf
C:Genetics:
A:Gene: dnaJ
A:Note: Induced upon tuberization
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
C:Keywords: molecular chaperone

Query Match 28.4%; Score 534; DB 2; Length 419;
Best Local Similarity 36.7%; Pred. No. 4.8e-37;
Matches 130; Conservative 68; Mismatches 112; Indels 44; Gaps 13;

QY 26 FYKILGVPRSAIKDIAKAYRKALQLPDRNPDDPQAEKFDLGAAYVLSDESKRQ 85
Db 14 YETILGVPRKTAAGDLKAYRKALKNHDPKGGD---PEKFKLAQAYVLSDESKREI 69
QY 86 YDTYGEGLKDGHOSSHG-----DIFSHFFG--DFGFMFGTTPROODRNIPRGSIDIVLD 138
Db 70 YDQYGEALKEGMMGGGGGMDPFDIFSSFFGSGPFG---GGGSSRRQRGRGDEVVHPL 126
QY 139 EYVLEEVYAGNF--VEVYRN-----KVARQAPGRKCN-CR-QEMRTT--QLGPRFQ 186
Db 127 KVSLEDLYNGTSKSLSLSRNVLCSCCKGSKGSGASMKSGCGSGKMTIRQLGSPMTQ 186
QY 187 MTQ-----EVV-----CDECPNVKLVNEERTLEVEIPGVDRGMETPFIGSEPH 231
Db 187 QMHPCNECKGTGEMINDKRCQCGKGEKVQEKVLEVVVEKMGQGRKITEFGADEA 246
QY 232 VDEPGLDPRIRIKVYKHPFERRQDLYTNVTVSLVESLVGFEMDITHLQHKVHI--SR 289
Db 247 PDVTGDIYVFLQOKHPRKRRKGDLEFVHTLSLTALCGFQFIIHLNDRQLITRQA 306
QY 290 DKITRPGAXXMKKGEGLPNFDNNINKSLITFDVDPKQDLTEAREGIKOLL 343
Db 307 GEVYKPFQKAIINDEGMPYGRPFMRGKLYIHFTVEFP-DTSLSPQCKNLEAVL 359

RESULT 6
JQ2142
chaperone ANO1 protein - Atriplex nummularia
C:Species: Atriplex nummularia
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 05-Dec-1997
C:Accession: JQ2142
R:Zhu, J.K.; Shi, J.; Bressan, R.A.; Hasegawa, P.M.
Plant Cell 5, 341-349, 1993
A:title: Expression of an Atriplex nummularia gene encoding a protein homologous to the
A:Reference number: JQ2142; MUID:93222693
A:Accession: JQ2142
A:Molecule type: mRNA
A:Residues: 1-417 <ZHU>
A:Cross-references: GB:I.09124
C:Comment: This protein is involved in protein folding and is observed as heat shock prc
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
C:Keywords: molecular chaperone

F:13-74/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 28.0%; Score 526.5; DB 2; Length 417;
Best Local Similarity 34.9%; Pred. No. 2e-36;
Matches 123; Conservative 70; Mismatches 118; Indels 41; Gaps 11;

QY 26 FYKILGVPRSAIKDIAKAYRKALQLPDRNPDDPQAEKFDLGAAYVLSDESKRQ 85
Db 14 YETILGVPRKTAAGDLKAYRKALKNHDPKGGD---PEKFKLAQAYVLSDESKREI 69
QY 86 YDTYGEGLKDGHOSSHG-----DIFSHFFGDFGFMFGTTPROODRNIPRGSIDIVLD 140
Db 70 YDQYGEALKEGMMGGGGGMDPFDIFSSFFGSGPFG---GGGSSRRQRGRGDEVVHPLK 127
QY 141 TLEEVYAGNFVEV--VRNKPVARQAPGR-----RKCN-CR--QEMRTTQLGPRPMT 188
Db 128 SLEDLFTGTTRKSLSLSRNVLCSCCKGSKGSGASMKSGCGGTGMKVSIRHLGSPMTQM 187
QY 189 Q-----EVV-----CDECPNVKLVNEERTLEVEIPGVDRGMETPFIGSEPH 233
Db 188 QHPCNECKGTGETINDKRCPCQCKGEKVQEKVLEVVVEKMGQGRKITEFGADAPD 247
QY 234 GEPGLDPRIRIKVYKHPFERRQDLYTNVTVSLVESLVGFEMDITHLQHKVHSRD--K 291
Db 248 TVTGDIVYVFLQOKHPRKRRKGDLEFVHTLSLTALCGRFVTLHLDGRQLIKSNLGE 307
QY 292 ITRPGAXXMKKGEGLPNFDNNINKSLITFDVDPKQDLTEAREGIKOLL 343
Db 308 VKPQGFKAIDEGMPYGRPFMKGMVYHFTVEFP-DSLNPQGVKLEAIL 358

RESULT 7
S33312
dnaJ protein - leek (fragment)
C:Species: Allium porrum (leek)
C:Date: 06-Jan-1995 #sequence_revision 01-Sep-1995 #text_change 20-Aug-1999
C:Accession: S33312; S31387
R:Bessoule, J.J.
FEBS Lett. 323, 51-54, 1993
A:title: Occurrence and sequence of a DnaJ protein in plant (Allium porrum) epidermal
A:Reference number: S33312; MUID:93265942
A:Accession: S33312
A:Molecule type: mRNA
A:Residues: 1-397 <BES>
A:Cross-references: EMBL:X69436; NID:g16086; PIDN:CAA49211.1; PID:g16087
C:Genetics:
A:Gene: dnaJ
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
F:1-53/Domain: dnaJ amino-terminal homology (fragment) <DNJ>

Query Match 27.6%; Score 519.5; DB 2; Length 397;
Best Local Similarity 35.6%; Pred. No. 7.2e-36;
Matches 124; Conservative 66; Mismatches 117; Indels 41; Gaps 9;

QY 34 RSASIKDIAKAYRKALQLPDRNPDDPQAEKFDLGAAYVLSDESKRQYTYEEG 93
Db 1 KNASPDLLKRAYRKALKNHDPKGGD---PEKFKLAQAYVLSDESKREIYDQYEDA 56
QY 94 LKDGHOSSHG-----DIFSHFFGDFGFMFGTTPROODRNIPRGSIDIVLDLEEVYAG 148
Db 57 LKEGMMGGGGGMDPFDIFSSFFGSGGSSR--GRQRGRGDEVVHPLVSLLELYNG 114
QY 149 NFVYVANKPY-----AQAPGRKRCNCR--QEMRTTQLGPRPMTQ----- 190
Db 115 TSKLISLRNVICSKGSKGSGASMKSGCGSGKVSIRQLGPRGMIDQMHPKNDCK 174
QY 190 ---EVV-----CDECPNVKLVNEERTLEVEIPGVDRGMETPFIGSEPHVDEGPRDLR 241
Db 175 GTGEMINDKRCPLCKGEKVQEKVLEVVVEKMGQGRKITEFGADEAPDVTGDIYV 234
QY 242 RIKVYKHPFERRQDLYTNVTVSLVESLVGFEMDITHLQHKVHSRD--KITRPGAXX 299

Db 235 VLQKREHPKFGKDDLFYKRTLSLTALCFQVLTHTLDRQLLSNPEVVKPQDFK 294
 QY 300 WKKEGLPNFDNNNKISLITFDVDPKLEGEAREGITKQLKQGS 347
 Db 295 AINDEGHPMYORPPMKGLYIQFLVDFP-DSLTPDQCKVLESVLPKSA 341

RESULT 8 S34630

dnaJ protein homolog - human
 C:Species: Homo sapiens (man)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S34630; S34631
 R:Chelliah, A.; Davis, A.; Mohanakumar, T.
 Biochim. Biophys. Acta 1174, 111-113, 1993
 A>Title: Cloning of a unique human homologue of the Escherichia coli DnaJ heat shock pro
 A:Reference number: S34630; MUID:93326629
 A:Accession: S34630
 A:Molecule type: mRNA
 A:Residues: 1-397 <CHE>
 A:Cross-references: EMBL:L08069; NID:9306713; PIDN:AAC37517.1; PID:9306714
 A:Experimental source: umbilical vein endothelium
 R:Oh, S.; Iwahori, A.; Kato, S.
 Biochim. Biophys. Acta 1174, 114-116, 1993
 A>Title: Human cDNA encoding DnaJ protein homologue.
 A:Reference number: S34631; MUID:93326630
 A:Accession: S34631
 A:Molecule type: mRNA
 A:Residues: 1-273, 'H', 275-397 <TM>
 A:Cross-references: EMBL:DJ3388; NID:9219587; PIDN:BAA02656.1; PID:d1003160; PID:g219588
 A:Experimental source: fibrosarcoma HT-1080
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:6-68/Domain: dnaJ amino-terminal homology <DMJ>

Query Match 27.4%; Score 515.5; DB 2; Length 397;
 Best Local Similarity 34.5%; Pred. No. 1.6e-35;
 Matches 133; Conservative 66; Mismatches 117; Indels 69; Gaps 13;

QY 26 FYKILGVPKRSIKDKAYRKALQLHPDRNDPQAEKFDGLGAAYVLSDESKRQ 85
 Db 7 YIVLVGKPRATQELKATRKALTKHPDKPNE--GKEKQISQAYEVLSDAKKREL 63
 QY 86 YDTYGEGLKDGHO---SSHGDIFFSHFGDFGFMFGTTPROQDRNIPRGSIIIVDLEVT 141
 Db 64 YDKGEGQAIKEGAGGAGGFGSPMDIFDMFEG-----GGGMQRER--RKNVNHQLSVT 114
 QY 142 LEEVYAGNFEVYVNRKPV---ARQAPGR--KC--NCR---QEMRTTQLGGRFQMTQ 189
 Db 115 LEDLVNGATRKALKQKNVLCIDCKEGRGKGAVECCPNCRGTMQIRIHQIGGMVQQLQ 174
 QY 190 EYV-----ODECPNVLVNERLLEVEIEGVDGMEYPTIGGGEHVDS 224
 Db 175 SVMEQCGHGERISPRDRCSGKGRIVREKTLLEVIDGMDGKQITFFGGDEPGU 234
 QY 235 EPGDLFRKIVVHPYIFERKGLDLYNVTVLSVLESGFEMDTTHDGHKVNHSR--DKI 292
 Db 235 EPEDDITIVLDQKHAVFTRRGEDLFMCMDIQVLEALCGQKRPSTLDNRIVTISPGQI 294
 QY 293 TRPGAXMKKKGEGLPNFDNNNKISLITFDVDPK-----EQLTEAREGIT 340
 Db 295 VKHGDKICVNLNEGMPYRRPYREKGRILIEFKVNPENGFLSPDKLSLEKLLPERKE-VE 353
 QY 341 OLLKGGYQKV-----YNG 354
 Db 354 ETDEMDQVELVDFPNOERRRHNG 378

RESULT 9
 T01643
 DnaJ protein homolog ZMDJ1 - maize
 C:Species: Zea mays (maize)

C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 20-Aug-1999
 C:Accession: T01643
 R:Baszczynski, C.L.; Barbour, E.; Zeka, B.; Maddock, S.E.; Swenson, J.L.
 Maylica 42, 189-201, 1997
 A>Title: Characterization of a genomic clone for a maize DnaJ-related gene, zmdj1, an
 A:Reference number: Z14364
 A:Accession: T01643
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-419 <BAS>
 A:Cross-references: EMBL:AF053468; NID:92984708; PIDN:AAC08009.1; PID:g2984709
 C:Genetics: mdj1
 A:Gene: mdj1
 A:Introns: 50/3; 105/1; 153/2; 245/3; 312/1
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 C:Keywords: heat shock; molecular chaperone

Query Match 27.3%; Score 515; DB 2; Length 419;
 Best Local Similarity 35.8%; Pred. No. 1.9e-35;
 Matches 125; Conservative 61; Mismatches 115; Indels 48; Gaps 10;

QY 26 FYKILGVPKRSIKDKAYRKALQLHPDRNDPQAEKFDGLGAAYVLSDESKRQ 85
 Db 14 YVEILGVPKRSASQDDLKAYRKALKHNPDKGD---PEKFKELAQAYEVLSDEPEKREI 69
 QY 86 YDTYGEGLKDGHO---GHSN---GDTSHPFG-DFGFMFGTTPROQDRNIPRGSIIIVDLE 139
 Db 70 YDQYGEDALKEGKGGGSHDPPDIFSFEGPFG---GGGSSRRRORRGEDVNHPRK 126
 QY 140 VTELEVYAGNFEVYVNRKPV-----ARQAPGRKCRORQMRRTQLGGRF 185
 Db 127 VSEDLVNGSKSLISLRNVICSKCKGSKGASARCCGCGSSGKAVYIR--QLSPSKI 184
 QY 186 QMTQEVNCDCEPNV-----KLVNERLTLEIPEGVDRGMEYPTIGEGE 229
 Db 185 IQMCOQ-PCNECKCTGSEINEKDRPCCKGKGVYQEKVLEVHEKQJHNOQKITTEPGAD 243
 QY 230 PNVDPGEGDREFRIKIVKHPFERRGDDLYTNTVLSVLSVGFEMDTTHDGHKVNHSR 289
 Db 244 EAPDVTGDIIVFLVQKDKSKFRKGEDLYEHTLSLTALCFQVLTHTLDRQLLSN 303
 QY 290 D-KITRPGAXMKKKGEGLPNFDNNNKISLITFDVDPKLEGEAREGIT 336
 Db 304 DPGEVVKPPOQKAINDEGMPYORPPMKGLYIHFTVEPDSLAPDQCK 352

RESULT 10
 S26703
 dnaJ protein homolog YDJ1 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: dnaJ protein homolog MAS5; protein N2418; protein YNL064c; protein
 C:Species: Saccharomyces cerevisiae
 C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 20-Aug-1999
 C:Accession: S26703; A39659; S58714; S62992; S17250
 R:Attencio, D.P.; Yaffe, M.P.
 Mol. Cell. Biol. 12, 283-291, 1992
 A>Title: MAS5, a yeast homolog of DnaJ involved in mitochondrial protein import.
 A:Reference number: S26703; MUID:92107179
 A:Accession: S26703
 A:Molecule type: DNA
 A:Residues: 1-409 <ATE>
 A:Cross-references: GB:S74758; NID:9241522; PIDN:AAB20771.1; PID:g241523
 R:Caplan, A.J.; Douglas, M.G.
 J. Cell Biol. 114, 609-621, 1991
 A>Title: Characterization of YDJ1: a yeast homologue of the bacterial dnaJ protein.
 A:Reference number: A39659; MUID:91332099
 A:Accession: A39659
 A:Molecule type: DNA
 A:Residues: 1-409 <CAP>
 A:Cross-references: GB:X56560; NID:94810; PIDN:CAA39910.1; PID:g4811
 R:Bergez, P.; Dolignon, F.; Crouzet, M.
 Yeast 11, 967-974, 1995
 A>Title: The sequence of a 44 420 bp fragment located on the left arm of chromosome X

Db 246 HVLPGMKHNDKITEKGDQSDPDGPDGVYIVIOQKDHDFKRDGDLHMTKLLSNEA 305
 QY 270 LVGFEMDITLHGHKVNHS--RDKITRPGAXXMKKGEGLPNDNNITGSLITFDVDP 327
 Db 306 LCGYFELIKHLDGHPVLVSSKQGVYIKGVYIRGVGKMPKKYPELKNLFVEFEVEP 365
 QY 328 KEOLTEAR 336
 Db 366 KEHFLDDEK 374

RESULT 13

E70361
 Chaperone DnaJ - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Aug-1999
 C:Accession: E70361
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666
 A:Accession: E70361
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-376 <AOF>
 A:Cross-references: GB:AE000703; NID:92983287; PIDN:ANC06881.1; PID:92983289; GB:AE00065
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: dnaJ2
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:8-72/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 26.0%; Score 490; DB 2; Length 376;
 Best Local Similarity 34.5%; Pred. No.2e-33;

Matches 128; Conservative 61; Mismatches 118; Indels 64; Gaps 15;

QY 24 RDEYKILGVPRASIKDIKKAYRKLALQHPDRNPDPQAEKFDLGAAYEVLSDSEK 83
 Db 7 KDYEELIGVPRNASQOEIKKAYRRLVRYHPD-ICKPECEKEKKEINAYOVLSDPEK 65
 QY 84 KQDVTYGEELKNGHSHG-----DIEHFEG-DEGFNF---GGTPRO 123
 Db 66 KLYDMT-----GHAFFEAGAOQVETTELPIREILREFFDFDGSIFERATGRRRA 118
 QY 124 QDRNIPRGSDDIIVLEVTLEEVYAGNFVEVVRKPYARQAPG-----KRC-NCRQE 174
 Db 119 RRRRSVKGEIVVPELTLEAEKGTTPLEVEREVPSCAGGTGDESKSRCPPTGGR 178
 QY 175 MRTLQGPGRFQMTQEVVCECPNVKLVN-----ERTLEVEIEPGRDGM 221
 Db 179 GETVQ-GNMFQYRQ--TCPTCGEGVIYENCHACTGRGYGLVKETLIKVPPIPGVADGS 235
 QY 222 YPIFGGEPRV---DDEPGLRIRIVVKHPIFERGGDLTYVTVSLVESLVGFEMDIT 278
 Db 236 --LVVGGKAGAYGPGGLYIVVYKPKIFERKGDLYDVVNITTEAVYLGTEVEP 293
 QY 279 HLDGKHVHSRDKITRPGAXXMKKGEGLPNDNNITGSLITFDVDPK---EOLTE 334
 Db 294 TLDEKLVKVIIPPGTEGELIKVPKGMRLKSG-RGDLIVVNHIDYKIGVLSKLDD 352
 QY 335 AREGILQQLKQ 345
 Db 353 GRK-VEELKK 362

RESULT 14

T22648
 Hypothetical protein F54D5.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22648

R:Coles, L.; Matthews, L.
 submitted to the EMBL Data Library, October 1995
 A:Reference number: Z19593
 A:Accession: T22648
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-331 <NTL>
 A:Cross-references: EMBL:266513; NID:e1062016; PIDN:CAA91334.1; GSPDB:GN00020; CESP:F
 A:Experimental source: clone F54D5
 C:Genetics:
 A:Gene: CESP:F54D5.8
 A:Map position: 2
 A:Introns: 46/3; 289/2

Query Match 25.0%; Score 470.5; DB 2; Length 331;
 Best Local Similarity 32.4%; Pred. No.7.1e-32;
 Matches 119; Conservative 55; Mismatches 102; Indels 91; Gaps 9;

QY 23 GRDFYKILGVPRASIKDIKKAYRKLALQHPDRNPDPQAEKFDLGAAYEVLSDSEK 82
 Db 2 KDYIKVLGISKATDDELKAYRKALKYHPDKN-KEAGAEKKFEIAEAYDVLSDDKK 60
 QY 83 RKQDVTYGEELKNGD-----HQSNGD---IFSHFGD----- 113
 Db 61 KKIYDQFGEGLKEGGGAGGGGGGMHYEFRRDPNNITSSFFGSGDPGAGGPGMFDLG 120
 QY 113 -----FGFMGTPRQDRNIPRGSDDIIVLEVTLEEVYAGNFE 152
 Db 121 GGAGGPFMEFMNOGMDNDFGGMHOGGRGHRQDPA--VLHDLVSLEVLKNG---- 174
 QY 153 VVNRKPYARQAPGRKRCNCRQEMRTQLGPRPOMQEVVCECPNVKLVNEERTLEVEI 212
 Db 174 -----TTK-----KKITIKVYNDNAORL-----EDKVLVTI 201
 QY 213 EPGVRDGMEXPIFGEPRHYDGPDLRIRIVVKHPIFERGGDLTYVTVSLVESLVG 272
 Db 202 KPGMKSGTITFPKEGDQHPNRTPADIVFYIKDKPHPKFRGSDIKRREKISTLSALG 261
 QY 273 FEEDITLHGHKVNHSRDKITRPGAXXMKKGEGLPNDNNITGSLITFDVDPREOLT 332
 Db 262 LDMITPLTGADYRLQLNLNVIRKPTRRLTGLGKLPNPKPSHRCDLITFEVDFP-SQLN 320
 QY 333 EEARREGI 339
 Db 321 PPGREVI 327

RESULT 15

B47042
 heat shock protein dnaJ - Brucella ovis

C:Species: Brucella ovis
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999

C:Accession: B47042
 R:Cellier, M.F.; Teyssier, J.; Nicolas, M.; Liautard, J.P.; Marti, J.; Sri Widada, J.

J. Bacteriol. 174, 8036-8042, 1992

A:Title: Cloning and characterization of the Brucella ovis heat shock protein dnaJ fu

A:Reference number: A47042; MUID:93094135

A:Accession: B47042

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-375 <CEL>

A:Cross-references: EMBL:M95799; NID:9144114; PIDN:ACG36133.1; PID:9144116

C:Note: sequence extracted from NCBI backbone (NCBIN:119966, NCBI:P:119968)

C:Genetics:

A:Gene: dnaJ

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

C:Keywords: DNA replication; heat shock; molecular chaperone

F:4-69/Domain: dnaJ amino-terminal homology <DNJ>

F:76-108/Region: G/F motif

F:146-153/Region: CXXCXGXG repeat

F:163-170/Region: CXXCXGXG repeat

F:185-192/Region: CXXCXGXG repeat

F:199-206/Region: CXXCXGXG repeat

Query Match 24.9%; Score 469; DB 2; Length 375;
Best Local Similarity 32.8%; Pred. No. 1.1e-31;
Matches 118; Conservative 59; Mismatches 129; Indels 54; Gaps 12;

```
QY 25 DEYKIIIVPRPSASIKDKKAYRKLALQHPDRPDDPQAGKQFODLGAAYEVLSDSERRK 84
   ||: ||| ||: ||: ||: ||| ||: ||| ||| ||: ||: ||| ||: ||: ||
Db 4 DYEAAGVTRTADDKTLKAAFKRLAQYHPDRPDDEAERKFEIGEAETLKDPOKRA 63
QY 85 QYDTYGEEGIKDGHQ-----SSHGDIFSHEFGDFGFMFGGTPRQODRNIPRSD 133
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 64 AYDRFGHAFAFENGGMGGFGNGFAGAGGFADIFEDIFGE---MMCG--RRRRGAERGAD 118
QY 134 IIVDLEVTLEEVYAGNFEVY-----VRNKP-----VARQAPGRKRCNCRQEM 175
   : : ||| ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 119 LRYNMEVTLEEAAYAGKTAQIRVPTSLTCDECSGSAKPGSQPTTCTWCSGSGRCRAAGF 178
QY 176 ----RTQLGPRGRQMTQAEVVCDECPNVKLYNEERTLEVEI-----EPGVROGMEYPT 225
   || ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 179 FSYERTCPGCGNGRGOI IKD--PCKKCHGQGVRDGRSLSVNIPAVSRTPEPRIR-----LA 231
QY 226 GEGEPHYDGEF--GDLRPRIKVVKHPIFERRRDDLYTNVTVSLVESLVGFMDITHLDGK 284
   ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 232 GEGEAGLRGSPAGDLYTLLSVKPRHEFFQRDGADLYCKVPI SMTTAALGGQFEVSTLDGTQ 291
QY 285 VHSIRDKITRPGAXXWKKGEGLPNFDNNNIGSLIITFDVDFPKEQLTTEARBGIKQLK 344
   : : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 292 TRVKVPEGTQNGKQFRLKKGMPGL--RQSVTGLDYIQIDIEFP-QNLSKRQRELLERFEK 349
```

Search completed: June 30, 2000, 09:58:19
Job time: 939 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:13:20 ; Search time 16.52 Seconds
(without alignments)
312.869 Million cell updates/sec

Title: US-09-501-714-1

Perfect score: 1883
Sequence: 1 MAPONLSTFCLLLYLIGAV.....IKOLKQSGSVKYNGLQGY 358

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PCRTUS_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/beckfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1879	99.8	358	2	US-08-868-288A-1
2	1879	99.8	358	3	US-09-235-373-1
3	1879	99.8	358	3	US-09-388-993-1
4	515.5	27.4	397	2	US-08-868-288A-5
5	515.5	27.4	397	3	US-09-235-373-5
6	515.5	27.4	397	3	US-09-388-993-5
7	515	27.3	419	2	US-08-868-417-3
8	448.5	23.8	348	2	US-08-868-417-1
9	437	23.2	340	2	US-08-974-546-5
10	345	18.3	330	2	US-08-472-534-6
11	257	13.6	330	2	US-08-868-288A-3
12	257	13.6	330	2	US-09-235-373-3
13	257	13.6	330	3	US-09-388-993-3
14	231.5	12.3	277	2	US-08-868-288A-7
15	231.5	12.3	277	3	US-09-235-373-7
16	231.5	12.3	277	3	US-09-388-993-7
17	231.5	12.3	351	2	US-08-868-288A-6
18	231.5	12.3	351	3	US-09-235-373-6
19	231.5	12.3	351	3	US-09-388-993-6
20	194	10.3	438	2	US-08-897-340-34
21	192	10.2	484	1	US-08-879-260-4
22	154.5	8.2	51	1	US-08-346-849-12
23	154.5	8.2	51	2	US-08-293-284A-12
24	154	8.2	52	1	US-08-346-849-8
25	154	8.2	52	2	US-08-293-284A-8
26	151	8.0	52	1	US-08-346-849-10
27	151	8.0	52	1	US-08-293-284A-10
28	145	7.7	52	1	US-08-346-849-9
29	145	7.7	52	2	US-08-293-284A-9

30	121.5	6.5	49	1	US-08-346-849-11	Sequence 11, Appl
31	121.5	6.5	49	2	US-08-293-284A-11	Sequence 11, Appl
32	100.5	5.3	550	2	US-08-844-058-2	Sequence 2, Appl
33	99	5.3	433	1	US-08-346-849-2	Sequence 2, Appl
34	99	5.3	433	2	US-08-293-284A-2	Sequence 2, Appl
35	97.5	5.2	51	1	US-08-346-849-14	Sequence 14, Appl
36	97.5	5.2	51	2	US-08-293-284A-14	Sequence 14, Appl
37	95	5.0	56	1	US-08-346-849-28	Sequence 28, Appl
38	95	5.0	56	2	US-08-293-284A-28	Sequence 28, Appl
39	89	4.7	3135	1	US-08-323-1708-2	Sequence 2, Appl
40	86	4.6	459	2	US-08-481-6588-2	Sequence 2, Appl
41	86	4.6	459	2	US-08-477-504A-2	Sequence 2, Appl
42	86	4.6	459	2	US-08-486-756A-2	Sequence 2, Appl
43	86	4.6	459	2	US-08-485-8628-2	Sequence 2, Appl
44	86	4.6	459	3	US-08-787-739-2	Sequence 2, Appl
45	86	4.6	1482	1	US-08-026-138E-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-08-868-288A-1
: Sequence 1, Application US/08868288A
: Patent No. 5922567
: GENERAL INFORMATION:
: APPLICANT: Au-Young, Janice
: APPLICANT: Lal, Preethi
: APPLICANT: Bandman, Olga
: TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESS: Inocyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: Fastseq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/868,288A
: FILING DATE: June 3, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0309 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 358 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: SYNORAB01
: CLONE: 136466
: US-08-868-288A-1

Query Match 99.8%; Score 1879; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.3e-199;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
|||||
Db 1 MAPQNSTCCLLLLYLIGAVIAGRDFFYKILGVRASISINDIKAKYKRLAQLHPDNPDP 60
QY 61 POAOKFODLGAAYEVLSDSEKRRKQYDTYGEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
Db 61 POAOKFODLGAAYEVLSDSEKRRKQYDTYGEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
QY 121 PROQDNNIPRGSIDIYDLVLTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTOL 180
Db 121 PROQDNNIPRGSIDIYDLVLTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTOL 180
QY 181 GGRFQMTQEVVDCDECPNKLVEERTLEVEIEPGVRDMEYPFIGEGRPHVDGEPDGLR 240
Db 181 GGRFQMTQEVVDCDECPNKLVEERTLEVEIEPGVRDMEYPFIGEGRPHVDGEPDGLR 240
QY 241 FRKVVKHPFERRGDDLYTNVTVSLVESLVGFEMDITLHDGKHVHSRDKITRPGAXXW 300
Db 241 FRKVVKHPFERRGDDLYTNVTVSLVESLVGFEMDITLHDGKHVHSRDKITRPGAXXW 300
QY 301 KKGEGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIKOLLKQGSVQKVVYNGLOGY 358
Db 301 KKGEGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIKOLLKQGSVQKVVYNGLOGY 358
```

RESULT 2

US-09-235-373-1
; Sequence 1, Application US/09235373

```
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
US-09-235-373-1
```

Query Match 99.8%; Score 1879; DB 3; Length 358;
Best local Similarity 100.0%; Pred. No. 2.3e-199;

```
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPQNSTCCLLLLYLIGAVIAGRDFFYKILGVRASISINDIKAKYKRLAQLHPDNPDP 60
Db 1 MAPQNSTCCLLLLYLIGAVIAGRDFFYKILGVRASISINDIKAKYKRLAQLHPDNPDP 60
QY 61 POAOKFODLGAAYEVLSDSEKRRKQYDTYGEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
Db 61 POAOKFODLGAAYEVLSDSEKRRKQYDTYGEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
QY 121 PROQDNNIPRGSIDIYDLVLTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTOL 180
Db 121 PROQDNNIPRGSIDIYDLVLTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTOL 180
QY 181 GGRFQMTQEVVDCDECPNKLVEERTLEVEIEPGVRDMEYPFIGEGRPHVDGEPDGLR 240
Db 181 GGRFQMTQEVVDCDECPNKLVEERTLEVEIEPGVRDMEYPFIGEGRPHVDGEPDGLR 240
QY 241 FRKVVKHPFERRGDDLYTNVTVSLVESLVGFEMDITLHDGKHVHSRDKITRPGAXXW 300
Db 241 FRKVVKHPFERRGDDLYTNVTVSLVESLVGFEMDITLHDGKHVHSRDKITRPGAXXW 300
QY 301 KKGEGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIKOLLKQGSVQKVVYNGLOGY 358
Db 301 KKGEGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIKOLLKQGSVQKVVYNGLOGY 358
```

RESULT 3

US-09-388-993-1
; Sequence 1, Application US/09388993

```
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
US-09-388-993-1
```

Query Match 99.8%; Score 1879; DB 3; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.3e-199;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPONLSTFCLLLLYLGAIVAGDFFKILGVPRASASTIKDKKAYRKALQHLHPDRNPD 60
DB 1 MAPONLSTFCLLLLYLGAIVAGDFFKILGVPRASASTIKDKKAYRKALQHLHPDRNPD 60
QY 61 POAOKFQDLGAAYEVLSDSEKRRQYDTYGEGLKHQSHSGDIFSHFGDFGFMFGT 120
DB 61 POAOKFQDLGAAYEVLSDSEKRRQYDTYGEGLKHQSHSGDIFSHFGDFGFMFGT 120
QY 121 PROODRPIPGSDIIVLEVTLEEVYAGNFEVYVNRKPVARQAGKRCRCRQEMRTQ 180
DB 121 PROODRPIPGSDIIVLEVTLEEVYAGNFEVYVNRKPVARQAGKRCRCRQEMRTQ 180
QY 181 GPGFQMTQEVYDECENKLVNEERTLEVEIEPGVRDGMETPTIGEGEPHYDGPDLR 240
DB 181 GPGFQMTQEVYDECENKLVNEERTLEVEIEPGVRDGMETPTIGEGEPHYDGPDLR 240
QY 241 FRIVKVPHPFERRGDDLYTNVTVSLVESLVGFEMDITHLGHKVTLSRDKITRPGAXXW 300
DB 241 FRIVKVPHPFERRGDDLYTNVTVSLVESLVGFEMDITHLGHKVTLSRDKITRPGAXXW 300
QY 301 KKGGLNPFNDNNIKGSLITFDVDFPREQDLTEARREGIKOLKOGSVQKYNGLOGT 358
DB 301 KKGGLNPFNDNNIKGSLITFDVDFPREQDLTEARREGIKOLKOGSVQKYNGLOGT 358

RESULT 4
US-08-868-288A-5
Sequence 5, Application US/08868288A
Patent No. 5922567
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNAG-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,288A
FILING DATE: June 3, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank

CLONE: 306714
US-08-868-288A-5

Query Match 27.4%; Score 515.5; DB 2; Length 397;
Best Local Similarity 34.5%; Pred. No. 1.2e-48;
Matches 133; Conservative 66; Mismatches 117; Indels 69; Gaps 13;

QY 26 FYKILGVPRASASTIKDKKAYRKALQHLHPDRNPDPOAOKFQDLGAAYEVLSDSEKRRQ 85
DB 7 YVDVGVNPNATQELKAYRKALQHLHPDRNPDNPNPNE--GEKFKQISQAYEVLSDAKKREL 63
QY 86 YDTYGEGLKHQ-----SSHGDIFFSHFGDPRGFMFGTTPQODRNPGRSDIIVDLEVT 141
DB 64 YDKGGEQAIKKGAGGGGSPMDIFDMFG-----GGGRMQRER--RGNVNHQLSVT 114
QY 142 LEEVYAGNFEVYVNRKPV-----ARQAPGR--KC--NCR--OEMRTQLGPGFQMTQ 189
DB 115 LEDLYNGATRKALQKNTYICKCEGRGKKGAVCECPKRCGTGMQIRHQIGPGVQIQ 174
QY 190 EVV-----CDCEPNKLVNEERTLEVEIEPGVRDGMETPTIGEGEPHYDGP 234
DB 175 SVCEMCGHGERISPKDCKSCNGRIYERKKLIEVHIDKMGKDGQKITFEGEGDQEPGL 234
QY 235 EPGDLRFRIKXVKKHPFERRGDDLYTNVTVSLVESLVGFEMDITHLGHKVTLSRDKITRPGAXXW 292
DB 235 EPGDLRFRIKXVKKHPFERRGDDLYTNVTVSLVESLVGFEMDITHLGHKVTLSRDKITRPGAXXW 292
QY 293 TRPGAXXWKKKGEGNPFNDNNIKGSLITFDVDFPREQDLTEARREGIKOLKOGSVQKYNGLOGT 340
DB 293 VKHGDIKCVLNGMPTIRPREYKRLITFVKVNPENGFSPDKLSLEKLLPERKE-VE 353
QY 341 QLKOGSVQKY-----YNG 354
DB 354 ETDEMQUVELVDFPNQERRRHYNG 378

RESULT 5
US-09-235-373-5
Sequence 5, Application US/09235373
Patent No. 6001598
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNAG-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,373
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/868,288
FILING DATE: June 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 306714
;
US-09-235-373-5

Query Match      27.4%; Score 515.5; DB 3; Length 397;
Best Local Similarity 34.5%; Pred. No. 1.2e-48;
Matches 133; Conservative 66; Mismatches 117; Indels 69; Gaps 13;

QY 26 FYKILGVPRASISIKDIKAYRKIALQLHPDRNPDDPOAQEKFODLGAAYEVLSDSEKRRQ 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 YVDVLGVKPNATQDELKKAIRKALKYHPDKNPNE--GEKFOISQAYEVLSDAKKREL 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 YDTYGEGLKDGHQ-----SSHGDISSHFGDFGFMFGGTPROODRNIPRGSIIYDLEVT 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 64 YDKGEGQAIKEGAGGFGSPMDIFDMFG-----GGGRMRER--RKNVYHQLSVT 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 142 LEEVYAGNFVEYVRANKPV---ARQAFGR---KC--NCR---QEMRTQLGPGRFQMTQ 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 115 LEDLYNGATRKALKQKNVICDKCEGKGGKGAVECCPCNCRGTGMQIRIHQIGGVMQOIQ 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 190 EVY-----CDECPNKLVEERTLEVEIEPGRVDSGMEYPTIGEGEPHYDG 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 175 SYVMECGGHERISPKDRCKSCNGRKIVREKILLEVHIDGKMDGOKITFHGEGDQEBGL 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 EPGDLLRRIVVKNHPFERGGDLYTNVYSLVESLVGFEMDITLDGKHVHISR--DKI 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 235 EPDIIIVLDQKHAFVTRGEDLFMCMDIQLVEALCGFQKPISTLDNRTIYITSHPGOI 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 TRGAXXWKKGEGLPNFDNNNIGSLITFDVDFPK-----EQLTEARREGIK 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 295 VKHGDIKCVLNEGMPYRRPEYKGRLLIEFKVNPENGLSPDKLSLEKLLPERKE-VE 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 341 QLLKGSYQKV-----YNG 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 354 ETDEMDOVELVDPEDPNERRRRHYNG 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
; Sequence 5, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

;
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 306714
;
US-09-388-993-5

Query Match      27.4%; Score 515.5; DB 3; Length 397;
Best Local Similarity 34.5%; Pred. No. 1.2e-48;
Matches 133; Conservative 66; Mismatches 117; Indels 69; Gaps 13;

QY 26 FYKILGVPRASISIKDIKAYRKIALQLHPDRNPDDPOAQEKFODLGAAYEVLSDSEKRRQ 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 YVDVLGVKPNATQDELKKAIRKALKYHPDKNPNE--GEKFOISQAYEVLSDAKKREL 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 YDTYGEGLKDGHQ-----SSHGDISSHFGDFGFMFGGTPROODRNIPRGSIIYDLEVT 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 64 YDKGEGQAIKEGAGGFGSPMDIFDMFG-----GGGRMRER--RKNVYHQLSVT 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 142 LEEVYAGNFVEYVRANKPV---ARQAFGR---KC--NCR---QEMRTQLGPGRFQMTQ 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 115 LEDLYNGATRKALKQKNVICDKCEGKGGKGAVECCPCNCRGTGMQIRIHQIGGVMQOIQ 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 190 EVY-----CDECPNKLVEERTLEVEIEPGRVDSGMEYPTIGEGEPHYDG 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 175 SYVMECGGHERISPKDRCKSCNGRKIVREKILLEVHIDGKMDGOKITFHGEGDQEBGL 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 EPGDLLRRIVVKNHPFERGGDLYTNVYSLVESLVGFEMDITLDGKHVHISR--DKI 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 235 EPDIIIVLDQKHAFVTRGEDLFMCMDIQLVEALCGFQKPISTLDNRTIYITSHPGOI 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 TRGAXXWKKGEGLPNFDNNNIGSLITFDVDFPK-----EQLTEARREGIK 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 295 VKHGDIKCVLNEGMPYRRPEYKGRLLIEFKVNPENGLSPDKLSLEKLLPERKE-VE 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 341 QLLKGSYQKV-----YNG 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 354 ETDEMDOVELVDPEDPNERRRRHYNG 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
; Sequence 3, Application US/08686417
; Patent No. 5850018
; GENERAL INFORMATION:
; APPLICANT: Baszzyński, Chris
; APPLICANT: Barbour, Eric
; APPLICANT: Horowitz, Jeanine
; APPLICANT: Rosichan, Jeffrey L.
; TITLE OF INVENTION: AN EXPRESSION CONTROL SEQUENCE FOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

Query Match	27.3%	Score 515;	DB 2;	length 419;
Best Local Similarity	35.8%;	Pred. No. 1.5e-48;		
Matches 125;	Conservative 61;	Mismatches 115;	Indels 48;	Gaps 10

```

RESULT      8
US-08-974-546-1
; Sequence 1, Application US/08974546
; Patent No. 5945287
; GENERAL INFORMATION:

```

APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Puri
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

Query Match	23.8%	Score 448.5;	DB 2,	Length 348;
Best Local Similarity	30.9%	Pred. No. 2.6e-41;		
Matches 119; Conservative	61;	Mismatches 100;	Indels 105;	Gaps 11.

RESULT 9
US-08-974-546-5
; Sequence 5, Application US/08974546
; Patent No. 5945287
; GENERAL INFORMATION:

1 APPLICANT: Lal, Preeti
2 APPLICANT: Corley, Neil C.
3 APPLICANT: Shah, Purni
4 TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
5 NUMBER OF SEQUENCES: 6
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Incyte Pharmaceuticals, Inc.
8 STREET: 3174 Porter Dr.
9
10

Sat Jul 1 21:34:06 2000

us-09-501-714-1.ra1

Page 6

```

1 CITY: Palo Alto
2 STATE: CA
3 COUNTRY: USA
4 ZIP: 94304
5
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Diskette
8 COMPUTER: IBM Compatible
9 OPERATING SYSTEM: DOS
10 SOFTWARE: FASTSEQ for Windows Version 2.0
11
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/974,546
14 FILING DATE: Filed Herewith
15 PRIORITY APPLICATION NUMBER:
16 APPLICATION NUMBER:
17 FILING DATE:
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Billings, Lucy J.
20 REGISTRATION NUMBER: 36,749
21 REFERENCE/DOCKET NUMBER: PF-0428
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: 650-855-0555
24 TELEFAX: 650-845-4166
25 INFORMATION FOR SEQ ID NO: 5:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 340 amino acids
28 TYPE: amino acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 IMMEDIATE SOURCE:
32 LIBRARY: GenBank
33 CLONE: 1816452
34
35 US-08-974-546-5

```

Query Match	23.28;	Score 437;	DB 2;	Length 340;
Post Total	30.18;	Score 460;	DB 2;	Length 340;

```

0Y      23 GDEFKILGVPASATIKDKKRYKLTALOLDHNRDPDPOAEKODGAAYEVLSDSEK 82
      2 GDDYQUTGLANGASADEETKRAIYKQALRYHBDK-KEPGAEKFKETAEYVYDLSDPK 60
0Y      83 RKQDYTYGEBGLKDKHOSS-----HGD---YFHFPG---DFGFMFGTP 121
      1: | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61 REIFDRYGEGLKSGSGSGSGGANGTGSFYTHRGDPHAFEEFGGRNFDFFG--- 118
0Y      122 KOODNPIRGSDI-----IVDLEVTLEE 144
      11 : | | | | | | | | | | | | | | | | | | | | | |
Db      118 ---QNGEBGMDIDPFSGFPMGMGFTNVNFGRSRSAQEPARKODPVYHDLRVSTEE 174
0Y      145 VYAGFVEYVNRKKPARAPARKKRCNCQNEAKRTQOLGREGROMTQEVYVCECPVAKLYNE 204
      1: | : | | | | | | | | | | | | | | | | | | | | | |
Db      175 IYSG-----CTKKMKISHRLNP-----DKSIRNE 200
0Y      205 ERTLEVEIEPGRVDMEXEPTIGEGBPHVDCGDLRFRIKVKVRIPERRDDLYTNVY 264
      : : | : | : | : | | : | | : | | : | | : | | : | | : | | : |
Db      201 DKILIEVKKKKEBETKITTFREBQDQISNNIPALIVYVLDKPKINIRKROGSDYIYARI 260
0Y      265 SLVESLVGFEMDITHLDSHKVHISBRTITPRGAXXMKKGEGLPNFDNNNIKGSIIITFDY 344
      11 : | : | | : : | | | : | | | | | | | | | | | | | | | | | |
Db      261 SLREMLACCTVNPVLDGRTIPVFEKIDVIRGMRKRYGEGDLPKTPREKRGDIIIEFV 320
0Y      325 DEPKQOLTEARREGIKOLL 343
      11 | : : : : | : : | |
Db      321 IFP--ERIPQTSRTVLEOVL 338

```

RESULT 10
US-08-472-534-6
; Sequence 6, Application US/08472534
; Patent No. 5919620
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee

```

1  APPLICANT: Brodeur, Bernard R
2  APPLICANT: Martin, Denis
3  TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM
4  TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE
5  NUMBER OF SEQUENCES: 6
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: Fish & Neave
8  STREET: 1251 Avenue of the Americas
9  CITY: New York
10 STATE: New York
11 COUNTRY: United States of America
12 ZIP: 10020
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patent In Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/472,534
20 FILING DATE:
21 CLASSIFICATION: 424
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Haley Jr, James F
24 REGISTRATION NUMBER: 27,794
25 REFERENCE/DOCKET NUMBER: Biovac-2
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 212-596-9000
28 TELEFAX: 212-596-9090
29 TELEK: 14-8367
30 INFORMATION FOR SEQ ID NO: 6:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 352 amino acids
33 TYPE: amino acid
34 TOPOLOGY: linear
35 MOLECULE TYPE: protein
36 US-08-472-534-6

```

Query Match	18.3%	Score 345;	DB 2;	Length 352;
Post Local Classification	38.4%	Score 71;	DB 2;	Length 352;

QY	25	DEYKLTGPRASSTIDIKKAYRKALALOHPRNRDDPQAOEKFODLGAAYVUSDSKKR	84
Dd	5	EYDRLGYSKSNASADEIKKAYRKLSKTHPIDN-KEPEADEKTYEVEDEAYETTSDDDKA	63
QY	85	QYDTYGERGLDGHOSHSNG-----DIFSHFFGDFGFMFGGPRRODRNIPR	130
Dd	64	AYDDYGAAGANGGFGGAGCGFGFGAGCGFGFEDIFSSFFBG-----GGSR--NPNAPR	116
QY	131	-GSDIYDLEAVTLEBVYAGNFVEVVRNKPVARQAPGKRCNCRCQEMPTTOLGPRFOMQ	189
Dd	117	QGDIDLOYVNLTFEBALFTEKEVKKYH--REA-GCRTCN-----GSGAKPGTS	161
QY	190	EYVDCDECPNKLTV-----EEXTLE	209
Dd	162	PVYTGRCRGAGAVINVDIOTPLGMMRROYTCDCVGRKSEKIKYPCOTTCGTHGNEKQASHSV	221
QY	210	VEIEPFGVADGMEYEPFIGEBEPHYDGE-GLDRFRKIKVKNHIFERRRGDDLYTNWTVSYVE	268
Dd	222	VKIPAGVETGGQIQLAAGOGEGFNGFGYGDLYVYVSVAESDKPERBETTLIFYNLNLNFVQ	281
QY	269	SLVGFEMDITHLDSGHKVAHISDKITFRGAXXWKKGEGLPNFDNNNINKGSLITTFEDVDFPK	328
Dd	282	AALDPTADYDIPVHG-DVELVPECTQGRKKFRLSKCAPSLRGAV-GDYVYVNVVTP-	339
QY	329	EQLTEARREGIKQ	341
Dd	339	TGLNDRQKVALE	351

RESULT 11
US-08-868-288A-3


```

? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0309 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? INFORMATION FOR SEQ. ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 277 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: GenBank
? CLONE: 32470
? US-08-868-288A-7

```

Query Match	12.3%	Score 231.5;	DB 2;	Length 277;
Best Local Similarity	47.6%;	Pred. No. 1,76-17;		
Matches	50;	Conservative 14;	Mismatches 22;	Indels 19; Gaps 3
QY	26	FYKILVPSASAKDKIKKAYRRKLALQLHPDRPPDPQ AQEKFDLCAALEVUSDSEKKR	84	
		::::: ::: ::: ::: ::: :		
Dd	4	LYELIVPNSASADDKIKKAYRRKRALQWHPDKDKNKEFAEKEFEVAEALEVLSDKHRE	63	
QY	85	OYETYGEEGL-----KDGHGSHS-----DIESHFG	111	
			:::	
Dd	64	IYIRYRGEGITGTCTGPSAEAGSGGGPGFTTFNSPEEVFVEEFEG	108	

```

1      RESULT 15
2      US-09-235-373-7
3      ; Sequence 7, Application US/09235373
4      ; Patent No. 6001598
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: Au-Young, Janice
9      ; APPLICANT: Lal, Preeti
10     ; APPLICANT: Bandman, Olga
11     ; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
12     ; NUMBER OF SEQUENCES: 7
13     ; CORRESPONDENCE ADDRESS:
14     ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
15     ; STREET: 3174 Porter Drive
16     ; CITY: Palo Alto
17     ; STATE: CA
18     ; COUNTRY: USA
19     ; ZIP: 94304
20     ; COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Diskette
22     ; COMPUTER: IBM Compatible
23     ; OPERATING SYSTEM: DOS
24     ; SOFTWARE: FastSeq for Windows Version 2.0
25     ; CURRENT APPLICATION DATA:
26     ; APPLICATION NUMBER: US/09/235,373
27     ; FILING DATE:
28     ;
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: 08/866,288
31     ; FILING DATE: June 3, 1997
32     ; ATTORNEY/AGENT INFORMATION:
33     ; NAME: Billings, Lucy J.
34     ; REGISTRATION NUMBER: 36,749
35     ; REFERENCE/DOCKET NUMBER: PF-0309 US
36     ; TELECOMMUNICATION INFORMATION:
37     ; TELEPHONE: 415-855-0555
38     ; TELEFAX: 415-845-4166
39     ; INFORMATION FOR SEQ ID NO: 7:
40     ; SEQUENCE CHARACTERISTICS:
41     ; LENGTH: 277 amino acids
42     ; TYPE: amino acid
43     ; STRADEDNESS: single
44     ; TOPOLOGY: linear
45     ; IMMEDIATE SOURCE:

```


THIS PAGE BLANK (USPTO)

34	86	4.6	1482	1	R45944	Glutamic acid rece
----	----	-----	------	---	--------	--------------------

35	85.5	4.5	953	1	R75707	Recombinant cold-
36	85.5	4.5	953	1	R90923	F. <i>bidentis</i> pyruva
37	85	4.5	806	1	R27481	RP-III residual pr
38	85	4.5	1482	1	R44193	Rat NMDA receptor
39	83	4.4	589	1	P60303	Sequence encoded b
40	83	4.4	764	1	W26727	Yeast Msp1p protea
41	82	4.4	393	1	W61252	Sallycyclic acid ind
42	82	4.4	1195	1	R05530	High density lipop
43	82	4.4	1292	1	R05531	NMDA receptor chan
44	82	4.4	1456	1	R49042	Protease B. <i>Procta</i>
45	81.5	4.3	354	1	R35221	

ALIGNMENTS

RESULT		1
W94065	ID	W94065 standard; Protein: 358 AA.
W94065;	AC	
09-APR-1999	DT	(first entry)
Human DnaJ-like protein, HSPJ1.	DE	
DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia	KM	
immune disorder; inflammation; tissue damage; diabetes; wound healing;	KW	
chromosome mapping.	OS	
Homo sapiens.	OS	
key	FH	Location/Qualifiers
Misc_difference 298	FH	
/label= unknown	FT	
/note= "encoded by AAN"	FT	
Misc_difference 299	FT	
/label= unknown	FT	
/note= "encoded by TAN"	FT	

12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534

Query match 99.88; Score 1879; DB 1; Length 358;

Best Local Similarity 100.0%; Pred. No. 6.7e-184;

```
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MAPQNLSFCLLLYLIGAVIAGDFYKILGVPRSASIDIKKAYKKALQLHDPNDPDD 600
Db 1 MAPQNLSFCLLLYLIGAVIAGDFYKILGVPRSASIDIKKAYKKALQLHDPNDPDD 600

```

QY 61 PQAQEKFDLGAAYEVLSDSEKROYDTYGEGLKDGSHSGDIFSHFDFGFMFGCT 120
DB 61 PQAQEKFDLGAAYEVLSDSEKROYDTYGEGLKDGSHSGDIFSHFDFGFMFGCT 120
QY 121 PRODDRNIPRSDIITVLEVTLEEVYAGNFYVVRNKKVAPQAPGRKRCNCRQEMRTTOL 180
DB 121 PRODDRNIPRSDIITVLEVTLEEVYAGNFYVVRNKKVAPQAPGRKRCNCRQEMRTTOL 180
QY 181 GPGFQMTQEVVCECPNVKLVNEERTLEVEIEPGVBDGMXPRTIGSGEPHVDSEPDRL 240
DB 181 GPGFQMTQEVVCECPNVKLVNEERTLEVEIEPGVBDGMXPRTIGSGEPHVDSEPDRL 240
QY 241 FRIVVKNHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLDGHRKHSRDKITPRGAXXW 300
DB 241 FRIVVKNHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLDGHRKHSRDKITPRGAXXW 300
QY 301 KKGGLPNDNNNKGSLITTFDVPFKEQLTEBAREGKIQOLKQSGVQKVVNGLQGY 358
DB 301 KKGGLPNDNNNKGSLITTFDVPFKEQLTEBAREGKIQOLKQSGVQKVVNGLQGY 358

RESULT 2
W00936
ID W00936 standard; Protein; 419 AA.
AC W00936;
DT 11-NOV-1997 (first entry)
DE Maize DnaJ-related protein.
KW DnaJ-related protein; chaperone; protein folding; DNA replication;
KW translation; peptide translocation; ZmDJI; promoter; maize;
KW transgenic plant; insecticide; antifungal; fungicide;
KW crop protection.
OS Zea mays L. cv. B73.
PN W09705260-A2.
PD 13-FEB-1997.
PE 12-JUL-1996; U11676.
PR 26-JUL-1995; US-0015522.
PA (PION-) PIONEER HI-BRED INT INC.
PI Barbour E, Baszczyński C, Horowitz J, Rosichan JL;
DR N-PSDB; T84330.
PT Tobacco DnaJ-related gene transcription/translation regulatory
PT sequence, ZmDJI - is intermediate between constitutive and tissue
PT specific promoters, partic. for control of antifungal and
PT insecticide genes
PS Example 1; Fig 2A-B; 26pp; English.
CC This polypeptide sequence is encoded by the maize DnaJ-related
CC gene (see T84330). DnaJ-related proteins assist in chaperone-
CC mediated protein folding and provide cell viability at high
CC temperatures. They are also involved in DNA replication,
CC translation and peptide translocation across intracellular
CC membranes. Due to this wide range of functions, DnaJ has a wide
CC range of effectiveness and the gene's promoter sequence, ZmDJI
CC (see T84329), is effective in a wide range of tissues. The
CC claimed ZmDJI promoter can be used in claimed methods for control
CC of antifungal or insecticidal genes in transgenic plants.
SQ Sequence 419 AA.

Query Match 27.3%; Score 515; DB 1; Length 419;
Best Local Similarity 35.8%; Pred. No. 2.1e-44;
Matches 125; Conservative 61; Mismatches 115; Indels 48; Gaps 10;

```

```

DB 127 VSLIEDLYNGTSKRLSLRNVCCKGKSGASMRCPCCGSGAKVTIR--OLQPSMT 184
QY 186 QMTQEVVCECPNV-----KLVNEERTLEVEIEPGVBDGMXPRTIGSGEPHVDSEPDRL 229
DB 185 QMOCQ-PCNCKCKGTGSEINCKRCPCGKEKVIQOEKVLVEVHEKGMQHNQKITPFGEAD 243
QY 230 PHVDGEPGDRFRIRKVIKHPFERRGDDLYTNVTVSLVESLVGFEMDITHLDGHRKHSR 289
DB 244 EAPDTYVGDIVFYQKDSKFRKGGEDLYEHTLSLTFALCGFQVTLHDMRQLLINS 303
QY 290 D--KITRPGAXXMKKGLPNDNNNKGSLITTFDVPFKEQLTEBAREGKIQOLKQSGVQK 336
DB 304 DPEVYKPDQFKALNDEGMPIYQRPFMKGLYIHFVFEPPDLAPRQCK 352

```

```

RESULT 3
W83397
ID W83397 standard; Protein; 419 AA.
AC W83397;
DT 23-FEB-1999 (first entry)
DE Maize DnaJ clone protein sequence.
KW Maize; Zea mays; transcriptional promoter; control sequence; ZmDJI;
KW DnaJ clone; corn; insecticide; antifungal protein; transgenic plant;
KW insect resistance; European corn borer; fungal resistance;
KW Ostrinia nubilalis.
OS Zea mays.
PN US5850018-A.
PD 15-DEC-1998.
PE 26-JUL-1996; 686417.
PR 27-JUL-1995; US-0015522.
PR 26-JUL-1996; US-686417.
PA (PION-) PIONEER HI-BRED INT INC.
PI Barbour E, Baszczyński C, Horowitz J, Rosichan JL;
DR N-PSDB; V72905.
PT Maize ZmDJI promoter - and corresponding recombinant expression
PT systems and transgenic plants
PS Example 1; Fig 2; 16pp; English.
CC The present invention describes the maize ZmDJI control sequence.
CC described are: (1) a recombinant expression system comprising the maize
CC ZmDJI control sequence linked to a heterologous protein coding sequence;
CC (2) a monocotyledonous plant, plant part or plant cell modified to
CC contain the expression system; (3) DNA comprising fragments of the ZmDJI
CC control sequence which retain transcription-initiating activity and/or
CC the function of the leader sequence; and (4) a method to regulate the
CC expression of a gene under control of the ZmDJI control sequence in
CC plant cell, plant parts or plants, where the plant is a monocotyledon,
CC comprising modifying the cell, part or plant containing the gene under
CC control of the ZmDJI control sequence to contain the DNA of (3) or RNA
CC of the same nucleotide sequence. Expression systems as described above
CC in which the heterologous sequence encodes an insecticidal or antifungal
CC protein can be used to produce transgenic maize plants resistant to
CC insects (e.g. European corn borer) or fungi (e.g. Ostrinia nubilalis).
CC The control sequence comprises a constitutive promoter that directs
CC transgene expression in various plant tissues, including 11-week-old
CC leaf blades, leaf whorls, leaf collars, stalk pith, stalk nodes, roots
CC and kernels. The present sequence represents the protein
CC encoded by a maize DnaJ clone from the present invention.
SQ Sequence 419 AA.

```

```

Query Match 27.3%; Score 515; DB 1; Length 419;
Best Local Similarity 35.8%; Pred. No. 2.1e-44;
Matches 125; Conservative 61; Mismatches 115; Indels 48; Gaps 10;

```

[illegible]

```

OY 190 EVCDECPNVLVN-----EERTLE 209
DB 162 FVTGCRGAGAVINVDQTPGLMMRQVTCVCHGRGKEIKYPCCTGTCGHEQASHVH 221
OY 210 VEIRPGVADGMEYFPIGGEPRHNGEP-GDLRFRIKVKVKNHIFERRGDDLTNTVTSLVE 268
DB 222 VKIAGVETGGOIRLQAGGEGFNGPYGDLVYVVSVEASDKFEREGTTFIYNLNFVQ 281
OY 269 SLVGFEMDITLHDGKVIHISRDKITRPGAXXMKKGEGLPNDDNNINIGSLITFEDVDEPK 328
DB 282 AALDDYDIPYVHG-DVELVPEGTGTGKRLRSKAPSLRGAV-GDOYTVNVVTP- 339
OY 329 EQLTEAREGKIQ 341
DB 339 TGLNDRQKVALKE 351

```

```

RESULT 6
W98475 ID W98475 standard; Protein; 369 AA.
AC W98475;
AT 31-MAR-1999 (first entry)
DE H. pylori GHPO 885 protein.
KM GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KM peptic ulcer disease.
OS Helicobacter pylori.
PN W09843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PI A1-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
DR WPI: 98-542293/46.
DR N-PSDB: X14194.
DR New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PS infections and gastrointestinal diseases
PS Claim 8; Page 853-855; 2054pp; English.
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 369 AA;

```

```

Query Match 18.3%; Score 344.5; DB 1; Length 369;
Best Local Similarity 28.7%; Pred. No. 4,6e-27;
Matches 100; Conservative 64; Mismatches 143; Indels 41; Gaps 10;
OY 26 FYKILGVPKASDKIKKRYKRLALQLHPDRNDPQADKQDGLAATVLSDSKKRQ 85
DB 5 YFEILFEVHSHQETKSKYKRLALYHPDRNAGDEAEKREKRLNEATVLSDEKKRL 64
OY 86 YDTYGEGLKDKSHQSHGDIFFSHFGD-----FGFNGGTTPROODNINIRGSDIYD 137
DB 65 YDRYCKKGLNQA-GASQGD-FSDPFEDLGSFEDAFGFARSKRQKSIAP---DYLOT 119
OY 138 LEVTLAEVYAGNFVEY-VANKPVANQAPK-----RKCNCROEMRTTQ----- 180
DB 120 LELSKREAVFGCKRTIKVQYQVSCSDGTAKDKALETCQCNQNGQVFMQGFMSFAQ 179
OY 180 -----LGRGFRMQEVCDECPNKLIVNEERTLEVEIRPGVADGMEYFPIGGEPRHNGEP 234
DB 180 TCGAAGCGKATYKP---CQACKTYYLIKDEIDALIPGIDDPNRMVLKKNKGEYK 236
OY 235 EFGDLRFRIKVKVKNHIFERRGDDLTNTVTSLVEVGFEMDITLHDGKVIHISDKITR 294
DB 235 EFGDLRFRIKVKVKNHIFERRGDDLTNTVTSLVEVGFEMDITLHDGKVIHISDKITR 294

```

```

DB 237 KRGLDYLAEQVKEDEHKRRCBCLFIKAPVFTTIALGHTIKVPSLKGDELKIPRNAR 296
OY 295 PGAXXMKKGEGLPNFDDNNINIGSLITFEDVDFPREQITEAREGRIKOL 342
DB 297 DKQTFARNEGKVKH-PSSYRGLSILVELQYIYK-SLNRQOGLLEKL 342
RESULT 7
W94066 ID W94066 standard; Protein; 330 AA.
AC W94066;
AT 09-APR-1999 (first entry)
DE Human DnaJ-like protein, HSPJ2.
KM DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;
KM immune disorder; inflammation; tissue damage; diabetes; wound healing;
OS Homo sapiens.
FH Key Location/Qualifiers
FH Misc_difference 239
FT /label= unknown
FT /note= "encoded by GST"
FT Misc_difference 272
FT /label= unknown
FT /note= "encoded by GNG"
FT Misc_difference 290
FT /label= unknown
FT /note= "encoded by GNG"
FT Misc_difference 291
FT /label= unknown
FT /note= "encoded by NCC"
FT Misc_difference 293
FT /label= unknown
FT /note= "encoded by NNC"
FT Misc_difference 298
FT /label= unknown
FT /note= "encoded by NCA"
FT Misc_difference 301
FT /label= unknown
FT /note= "encoded by TTN"
FT Misc_difference 302
FT /label= unknown
FT /note= "encoded by NNA"
FT Misc_difference 305
FT /label= unknown
FT /note= "encoded by NCC"
FT Misc_difference 309
FT /label= unknown
FT /note= "encoded by NNA"
FT W0985509-A2.
PD 10-DEC-1998.
PF 02-JUN-1998; U11182.
PR 03-JUN-1997; US-868288.
PA (INCY-) INCYTE PHARM INC.
PI Au-Young J, Bandman O, Lal P;
DR WPI: 99-070259/06.
DR N-PSDB: X06100.
DR New nucleic acid encoding human DnaJ-like proteins - for diagnosis,
PT treatment and prevention of cancer, immune disorders and
PT inflammation
PS Claim 22; Fig 3A-D; 73pp; English.
CC This represents a human DnaJ-like protein, HSPJ2. The invention provides
CC two human DnaJ-like proteins which are heat shock proteins J1 and J2
CC (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host
CC cells containing a vector comprising the nucleic acids are used for the
CC production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used
CC to raise Ab, therapeutically and to screen for specific binding agents.
CC Antagonists are used to treat or prevent a wide variety of solid cancers,
CC leukaemia and lymphoma; immune disorders (typical of many disclosed are
CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
CC and parasitic infections) and inflammation. Agonists may be used to treat
CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound

```

AC Y11105;

FT	/label= unknown
FT	/note= "encoded by TGA "
FT	Misc_difference
FT	96
FT	/label= unknown
FT	/note= "unknown by TAG "
FT	Misc_difference
FT	102
FT	/label= unknown
FT	/note= "encoded by RGR "
FT	Misc_difference
FT	103
FT	/label= unknown
FT	/note= "encoded by RGT "
FT	Misc_difference
FT	106
FT	/label= unknown
FT	/note= "encoded by RAG "
FT	Misc_difference
FT	109
FT	/label= unknown
FT	/note= "encoded by TGA "
FT	Misc_difference
FT	110
FT	/label= unknown
FT	/note= "encoded by GYR "
FT	Misc_difference
FT	111
FT	/label= unknown
FT	/note= "encoded by NKC "
FT	Misc_difference
FT	112
FT	/label= unknown
FT	/note= "unknown by NKT "
FT	Misc_difference
FT	113
FT	/label= unknown
FT	/note= "encoded by TYN "
FT	Misc_difference
FT	114
FT	/label= unknown
FT	/note= "encoded by YTT "
FT	Misc_difference
FT	115
FT	/label= unknown
FT	/note= "unknown by YAA "
FT	Misc_difference
FT	116
FT	/label= unknown
FT	/note= "encoded by YTT "
FT	Misc_difference
FT	117
FT	/label= unknown
FT	/note= "encoded by TSA "
FT	Misc_difference
FT	118
FT	/label= unknown
FT	/note= "encoded by TGA "
FT	Misc_difference
FT	121
FT	/label= unknown
FT	/note= "unknown by TAA "
FT	Misc_difference
FT	128
FT	/label= unknown
FT	/note= "encoded by TYN "
FT	Misc_difference
FT	129
FT	/label= unknown
FT	/note= "encoded by ARR "
FT	Misc_difference
FT	130
FT	/label= unknown
FT	/note= "encoded by CYA "
FT	Misc_difference
FT	132
FT	/label= unknown
FT	/note= "encoded by CAY "
FT	Misc_difference
FT	134
FT	/label= unknown
FT	/note= "unknown by AKC "
FT	Misc_difference
FT	135
FT	/label= unknown
FT	/note= "encoded by YAA "
FT	Misc_difference
FT	136
FT	/label= unknown
FT	/note= "unknown by NAA "
FT	Misc_difference
FT	137
FT	/label= unknown
FT	/note= "encoded by YKT "
FT	Misc_difference
FT	139
FT	/label= unknown
FT	/note= "encoded by GRR "

```
FT Misc-difference 140 /label= unknown
FT /note= "encoded by RWI"
FT Misc-difference 141 /label= unknown
FT /note= "encoded by CAW"
FT Misc-difference 142 /label= unknown
FT /note= "encoded by WYC"
FT Misc-difference 143 /label= unknown
FT /note= "encoded by CNM"
FT Misc-difference 146 /label= unknown
FT /note= "encoded by NNN"
FT Misc-difference 147 /label= unknown
FT /note= "encoded by WKG"
FT Misc-difference 148 /label= unknown
FT Misc-difference 149 /label= "encoded by GTK"
FT Misc-difference 151 /label= "unknown"
FT /note= "encoded by SYC"
FT Misc-difference 155 /label= "unknown"
FT /note= "encoded by GNG"
FT /label= "unknown"
FT /note= "encoded by TGN"
FT WO9739122-A2.
PD 23-OCT-1997.
PF 11-APR-1997; U06042.
PR 12-APR-1996; US-631184.
PA (MURRO-) MURRO PHARM INC.
PI Theoharides TC;
DR WPI: 97-526459/48.
DR N-PSDB: T88058.
PT Human and murine secreted proteins - useful to research or treat
PT diseases or disorders related to their function
PS Claim 11; Pages 78-79; 140pp; English.
CC The present sequence is a human secreted protein, which may
CC have nutritional uses, or cytokine and cell
CC proliferation/differentiation, immune stimulating or suppressing,
CC haematopoiesis regulating, tissue growth, activin/inhibin,
CC chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activities.
CC It can also be used to research or treat diseases/disorders related
CC to its function.
CC The partial cDNA clone AP162 was 1st isolated from a human adult
CC placenta cDNA library. The partial cDNA clones AM931, AM610, AM340,
CC AM282, AK647, AK583, AK533 and AK296 were 1st isolated from a human
CC foetal kidney cDNA library. The partial cDNA clones H617 and B89
CC were 1st isolated from a human peripheral blood monocyte cell (Tnl
CC or Tn2) cDNA library. The partial cDNA clone AM191 was 1st isolated
CC from a human ovary (PA-1 teratocarcinoma) cDNA library. The partial
CC cDNA clones AT711, AT705 and AT319 were 1st isolated from a human
CC lymphocyte and dendritic cell cDNA library. The partial cDNA clones
CC AS34 and AS32 were 1st isolated from a human foetal brain cDNA
CC library. The partial cDNA clone AR260 was 1st isolated from a human
CC adult retina cDNA library. The partial cDNA clones K640 and K39
CC were 1st isolated from a murine bone marrow (stromal cell line
CC FCM-4) cDNA library.
SQ Sequence 159 AA;
```

```
Query Match 10.4%; Score 195.5; DB 1; Length 159;
Best Local Similarity 43.4%; Pred. No. 2.3e-12;
Matches 36; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
```

```
OY 3 PONTSPCLLLLVIGAVIAGRDFYKILGVPRASIKDIRKAYRKALQLHPDRNPDPQ 62
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 4 POSIFIAICILMTTELILASKSYDILGVPRKASERQIKKAFHKLAMKYPDKN-KSPD 62
```

```
OY 63 AOEKFDUGAAYEVLSDSEKRRQ 85
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 63 AEAKFREIAEAYETLSDANXTKR 85
```

```
RESULT 15
ID W59132
AC W59132;
DE Mus musculus Tnfr Interactor (MTI-3) protein.
KW serine protease; tub interactor; treatment; obesity; cachexia;
KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW neurodegenerative disease; Alzheimer's disease; drug screening;
KW Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW amyloidrophic lateral sclerosis; spinocerebellar degeneration.
OS Mus musculus.
PN WO9812302-A1.
PD 26-MAR-1998.
PF 05-SEP-1997; U15627.
PR 21-JUL-1997; US-897340.
PR 17-SEP-1996; US-715032.
PA (MILL-) MILLENNIUM PHARM INC.
PI Errata PR, Gimeno CJ;
DR WPI: 98-217246/19.
DR N-PSDB: V11858.
PT Tub interactor genes - used to develop products for the treatment
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT diabetes
PS Claim 28; Fig 4; 120pp; English.
CC The sequence is that encoding the Tnfr Interactor protein (MTI-3)
CC which is a putative serine protease. TI genes function
CC in biochemical pathways involved in weight control and
CC related disorders. The products can be used for treating
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC or a related disorder such as diabetes. The products can
CC also be used to modulate cell cycle progression and apoptosis.
CC They can be used for treating neurodegenerative diseases
CC which are characterised by apoptosis, including Alzheimer's
CC disease, Parkinson's disease, Huntington's chorea, amyloidrophic
CC lateral sclerosis or spinocerebellar degenerations. The
CC products can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 438 AA;
```

```
Query Match 10.3%; Score 194; DB 1; Length 438;
Best Local Similarity 38.8%; Pred. No. 1.4e-11;
Matches 45; Conservative 23; Mismatches 26; Indels 22; Gaps 4;
```

```
OY 24 RDEYKILGVPRASIKDIRKAYRKALQLHPDRNPDPQAEKFDUGAAYEVL 78
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 324 KDYKILGVPRASIKDIRKAYRKALQLHPDRNPDPQAEKFDUGAAYEVL 78
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 79 DSEKRYQDT---YGEGLADGHOSMGDIFSHF-----GDGFMFG 118
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 384 DPKKRTYDSQDLDEGMNMGDPDAN-NIFKAFGGGPGGFSFGPAGNYFOFG 438
```

Search completed: June 30, 2000, 09:57:22
Job time: 2723 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 11:53:59 ; Search time 49.41 Seconds
(without alignments)
3619,903 Million cell updates/sec

Title: US-09-501-714-2

Perfect score: 1376

Sequence: 1 TCTCACCAGGACTCGGACT.....TTTATTTCATATGCAGT 1376

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/2/lna/5C_COMB.seq:*
4: /cgn2_6/prodata/2/lna/5D_COMB.seq:*
5: /cgn2_6/prodata/2/lna/6_COMB.seq:*
6: /cgn2_6/prodata/2/lna/PCRUS_COMB.seq:*
7: /cgn2_6/prodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1374	99.9	1376	4	US-08-868-288A-2
2	1374	99.9	1376	5	US-09-235-373-2
3	1374	99.9	1376	5	US-09-388-993-2
4	91.4	6.6	1330	4	US-08-868-288A-4
5	91.4	6.6	1330	5	US-09-335-373-4
6	91.4	6.6	1330	5	US-09-388-993-4
7	65.2	4.7	2349	4	US-08-974-546-2
8	65	4.7	672	2	US-08-486-955A-6
9	59.4	4.3	7218	1	US-08-232-663-14
10	53.4	3.9	1515	1	US-08-221-816B-1
11	53.4	3.9	1687	1	US-08-143-219-26
12	48	3.5	1700	4	US-08-897-340-4
13	47.8	3.5	4320	4	US-08-472-534-4
14	44.8	3.3	1756	4	US-08-879-260-3
15	40.8	3.0	20235	2	US-07-642-73AC-3
16	39.6	2.9	3748	3	US-08-686-417-2
17	38.8	2.8	1860	6	US-08-331-644-3
18	38.8	2.8	1860	6	PCT-US93-04102-3
19	38.6	2.8	1055	2	US-08-618-464-1
20	38.6	2.8	1055	6	PCT-US95-04896-1
21	38.6	2.8	5852	1	US-07-867-106-2
22	38	2.8	2277	2	US-08-676-967-2
23	38	2.8	2277	2	US-08-676-974-2
24	38	2.8	2277	4	US-09-098-487-2
25	36.8	2.7	2064	1	US-08-343-428-1
26	36.8	2.7	22846	4	US-08-469-461-3
27	36.8	2.7	22846	5	US-07-890-609-3

C	28	36.6	2.7	1164	7	5240849-4	Patent No. 5240849
C	29	36.6	2.7	2749	7	5240849-1	Patent No. 5240849
	30	35	2.5	2326	3	US-08-231-193A-41	Sequence 41, Appl
	31	35	2.5	2326	4	US-08-486-273A-41	Sequence 41, Appl
	32	35	2.5	2326	5	US-08-480-474-41	Sequence 41, Appl
	33	35	2.5	3243	3	US-08-231-193A-44	Sequence 44, Appl
	34	35	2.5	3243	4	US-08-486-273A-44	Sequence 44, Appl
	35	35	2.5	3243	5	US-08-480-474-44	Sequence 44, Appl
	36	35	2.5	3698	3	US-08-231-193A-43	Sequence 43, Appl
	37	35	2.5	3698	4	US-08-486-273A-43	Sequence 43, Appl
	38	35	2.5	4002	3	US-08-480-474-43	Sequence 53, Appl
	39	35	2.5	4002	4	US-08-231-193A-53	Sequence 53, Appl
	40	35	2.5	4002	5	US-08-486-273A-53	Sequence 53, Appl
	41	35	2.5	4002	5	US-08-480-474-53	Sequence 53, Appl
	42	35	2.5	4017	3	US-08-231-193A-49	Sequence 49, Appl
	43	35	2.5	4017	4	US-08-486-273A-49	Sequence 49, Appl
	44	35	2.5	4017	5	US-08-480-474-49	Sequence 49, Appl
	45	35	2.5	4053	3	US-08-231-193A-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-08-868-288A-2
Sequence 2, Application US/08868288A
Patent No. 5922567
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868, 288A
FILING DATE: June 3, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1376 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SYNORAB01
CLONE: 136466
US-08-868-288A-2

Query Match 99.9%; Score 1374; DB 4; Length 1376;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCACCGGAGCTCGGAGCTCCCGGAAAGTGAGACCGGAGAGAGGGGGCTAGTAGTG 60
 DB 1 TCTCACCGGAGCTCGGAGCTCCCGGAAAGTGAGACCGGAGAGAGGGGGCTAGTAGTG 60
 QY 61 TCTCTGCGGAGCAGGAGAACCCCGCGCCCGCGGTGTAGGCGGCTCACAGGGCGG 120
 DB 61 TCTCTGCGGAGCAGGAGAACCCCGCGCCCGCGGTGTAGGCGGCTCACAGGGCGG 120
 QY 121 GTGGGCTGGGAGCGAGCGGCGGAGGAGGCTGTGAGAGGTGTGTGAACAGAGCC 180
 DB 121 GTGGGCTGGGAGCGAGCGGCGGAGGAGGCTGTGAGAGGTGTGTGAACAGAGCC 180
 QY 181 CGGAGACAGAGAACCCAGTGGCTCCGAGAACCTGAGCACCTTTTGGCTGTCTCTATAC 240
 DB 181 CGGAGACAGAGAACCCAGTGGCTCCGAGAACCTGAGCACCTTTTGGCTGTCTCTATAC 240
 QY 241 CTCAATCGGGGGGTGATGTCGGGAGAGATTTCTATAGATCTTGGGGGTGCTCGAAGT 300
 DB 241 CTCAATCGGGGGGTGATGTCGGGAGAGATTTCTATAGATCTTGGGGGTGCTCGAAGT 300
 QY 301 GCTCTATAAAGATATTTAAAGCCCTATAGAACTAGCCCTGACCTCATCCGAGC 360
 DB 301 GCTCTATAAAGATATTTAAAGCCCTATAGAACTAGCCCTGACCTCATCCGAGC 360
 QY 361 CGGAACCTGATGTCACAAAGCCAGAGAAATTCAGAGATCTGGTGTCTTATAG 420
 DB 361 CGGAACCTGATGTCACAAAGCCAGAGAAATTCAGAGATCTGGTGTCTTATAG 420
 QY 421 GTTCTGATGATGATGAGAAACGGAACAGACATCTTATGTTGAAGAGATTTAA 480
 DB 421 GTTCTGATGATGATGAGAAACGGAACAGACATCTTATGTTGAAGAGATTTAA 480
 QY 481 GATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 DB 481 GATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 QY 541 ATGTTTGGAGAAACCCCTGTCAGCAAGACAAATATTCAGAGAGATGATATAT 600
 DB 541 ATGTTTGGAGAAACCCCTGTCAGCAAGACAAATATTCAGAGAGATGATATAT 600
 QY 601 GTAGATCTAGAACTGATGATGATGATGATGATGATGATGATGATGATGATG 660
 DB 601 GTAGATCTAGAACTGATGATGATGATGATGATGATGATGATGATGATGATG 660
 QY 661 AACAAACCTGTGGCAAGGAGGCTCTGGCAAGGAGTGTGCGCAAGAGATG 720
 DB 661 AACAAACCTGTGGCAAGGAGGCTCTGGCAAGGAGTGTGCGCAAGAGATG 720
 QY 721 CGGACCAACGAGTGGGCGCTGGGCGCTTCCAAATGACCAAGAGGTGTGCGAG 780
 DB 721 CGGACCAACGAGTGGGCGCTGGGCGCTTCCAAATGACCAAGAGGTGTGCGAG 780
 QY 781 TGCCCTAATGCTAAATAGTGAAGAAAGCAAGCTGGAATAGAAATAGAGCTGG 840
 DB 781 TGCCCTAATGCTAAATAGTGAAGAAAGCAAGCTGGAATAGAAATAGAGCTGG 840
 QY 841 GTGAGAGAGCGGATGAGTACCCCTTATGGAAGGTGAGCTGATGATGAGGAG 900
 DB 841 GTGAGAGAGCGGATGAGTACCCCTTATGGAAGGTGAGCTGATGATGAGGAG 900
 QY 901 CCGGAGATTTAGGCTTCCGAATCAAGTGTGCAAGCACCATATTTAAAGAGAG 960
 DB 901 CCGGAGATTTAGGCTTCCGAATCAAGTGTGCAAGCACCATATTTAAAGAGAG 960
 QY 961 GATGATTTGTACAAATGTGACAGTCTATAGTTGAGTCACTGCTTGTGAGATG 1020
 DB 961 GATGATTTGTACAAATGTGACAGTCTATAGTTGAGTCACTGCTTGTGAGATG 1020
 QY 1021 GATATTTACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1080
 DB 1021 GATATTTACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1080

QY 1081 GGAGGAAATNTGGAAGAAAGGGGCTCCCACTTTGACAAACAATATCAAG 1140
 DB 1081 GGAGGAAATNTGGAAGAAAGGGGCTCCCACTTTGACAAACAATATCAAG 1140
 QY 1141 GGCCTTTGATTAATCACTTTTGTGATGATGATGATGATGATGATGATGATG 1200
 DB 1141 GGCCTTTGATTAATCACTTTTGTGATGATGATGATGATGATGATGATGATG 1200
 QY 1201 AGAGAGATTAACAGCTACTGAAACAAAGGTCAGTGCAGAGATTAATGAC 1260
 DB 1201 AGAGAGATTAACAGCTACTGAAACAAAGGTCAGTGCAGAGATTAATGAC 1260
 QY 1261 CAAGATATGAGAGTGAATTAATGAGTGTGATGATGATGATGATGATGATG 1320
 DB 1261 CAAGATATGAGAGTGAATTAATGAGTGTGATGATGATGATGATGATGATG 1320
 QY 1321 TATATCTGCAAGGTTTTTTTGTGTGTGTTTTGTTTTTATTTCAATATGCAAGT 1376
 DB 1321 TATATCTGCAAGGTTTTTTTGTGTGTGTTTTGTTTTTATTTCAATATGCAAGT 1376
 RESULT 2
 US-09-235-373-2
 ; Sequence 2, Application US/09235373
 ; Patent No. 6001598
 ; GENERAL INFORMATION:
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Lal, Preeti
 ; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/235,373
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/868,288
 ; FILING DATE: June 3, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0309 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1376 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: SYNORAB01
 ; CLONE: 136466
 ; US-09-235-373-2
 Query Match 99.9%; Score 1374; DB 5; Length 1376;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCACCGGAGCTCGGAGCTCCCGGAAAGTGAGACCGGAGAGAGGGGGCTAGTAGTG 60

|||||
Db 1 TCTACCGGGAGCTCGGGAGCTCCCGGGAAGTGCACCGGCAAGAGAGGGGGCTAGCTACTG 60
QY 61 TCTCTGCGAACCGAGGAGACCCCGCGCCCCCGGTGTGAGGCGCCCTCACAGGGCGG 120
Db 61 TCTCTGCGAACCGAGGAGACCCCGCGCCCCCGGTGTGAGGCGCCCTCACAGGGCGG 120
QY 121 GTGGGCTGGGAGCGGAGCGGCGGCGGAGAGGCTGTGAGGAGTGTGTGAACAGAGCC 180
Db 121 GTGGGCTGGGAGCGGAGCGGCGGCGGAGAGGCTGTGAGGAGTGTGTGAACAGAGCC 180
QY 181 CGGAGACAGAGAACCATGAGCTCCGAGAACCTGAGACCTTTTCTGTTGCTCTATAC 240
Db 181 CGGAGACAGAGAACCATGAGCTCCGAGAACCTGAGACCTTTTCTGTTGCTCTATAC 240
QY 241 CTCATCGGGCGGTGATGTCGCGAGAGATTCTATAGATCTTGGGGGTGCTCGAAGT 300
Db 241 CTCATCGGGCGGTGATGTCGCGAGATTCTATAGATCTTGGGGGTGCTCGAAGT 300
QY 301 GCGCTATTAAGATATTAAGGCGCTATAGAACTAGCCCTGCAAGCTTCATCCGAC 360
Db 301 GCGCTATTAAGATATTAAGGCGCTATAGAACTAGCCCTGCAAGCTTCATCCGAC 360
QY 361 CGGAAACCTGATGATCCACAAGCCGAGAGAAATTCAGGATCTGGGTGCTGTTATAG 420
Db 361 CGGAAACCTGATGATCCACAAGCCGAGAGAAATTCAGGATCTGGGTGCTGTTATAG 420
QY 421 GTTCTGTCAAGTAGTGAAGAAAGGAAACAGTACGATCTTATGTTGAAGAGATTAAA 480
Db 421 GTTCTGTCAAGTAGTGAAGAAAGGAAACAGTACGATCTTATGTTGAAGAGATTAAA 480
QY 481 GATGTCATCAGAGCTCCCATGAGACATTTTTCACACTCTTTGGGGATTTGGTTTC 540
Db 481 GATGTCATCAGAGCTCCCATGAGACATTTTTCACACTCTTTGGGGATTTGGTTTC 540
QY 541 ATGTTTGAGAGAACCCCTCGTCACAGACAGAAATTTCCAAAGAGAGATATTATT 600
Db 541 ATGTTTGAGAGAACCCCTCGTCACAGACAGAAATTTCCAAAGAGAGATATTATT 600
QY 601 GTAGATCTAGAGTCACTTTTGGAGAGATATATGACAGAAATTTTGTGAAGATTAGA 660
Db 601 GTAGATCTAGAGTCACTTTTGGAGAGATATATGACAGAAATTTTGTGAAGATTAGA 660
QY 661 AACAAACCTGTGGCAAGGCTCTGCGAAAGGAGATGTCATTTGCGCAAGAGATG 720
Db 661 AACAAACCTGTGGCAAGGCTCTGCGAAAGGAGATGTCATTTGCGCAAGAGATG 720
QY 721 CGGACACCCAGCTGGGCGCTTCCAAATGACCCGAGAGGTGTGTCGACGAA 780
Db 721 CGGACACCCAGCTGGGCGCTTCCAAATGACCCGAGAGGTGTGTCGACGAA 780
QY 781 TGCCCTAATGTCAAACTAGTGAATGAAGAAAGAGTGAAGTGAAGTGAAGTGAAG 840
Db 781 TGCCCTAATGTCAAACTAGTGAATGAAGAAAGAGTGAAGTGAAGTGAAGTGAAG 840
QY 841 GTGAGAACCGGATGAGTACCCCTTATTTGAGAGAGTGAAGTGAAGTGAAGTGAAG 900
Db 841 GTGAGAACCGGATGAGTACCCCTTATTTGAGAGAGTGAAGTGAAGTGAAGTGAAG 900
QY 901 CCTGAGATTTAGAGTTCCGAATCAAGTTGTCAAGCACCCATATTGTAAGAGAGAGA 960
Db 901 CCTGAGATTTAGAGTTCCGAATCAAGTTGTCAAGCACCCATATTGTAAGAGAGAGA 960
QY 961 GATGATTTGTACAAATGTGACAGTCTCATTTAGTCACTGCTGCTTGAAGTGAAG 1020
Db 961 GATGATTTGTACAAATGTGACAGTCTCATTTAGTCACTGCTGCTTGAAGTGAAG 1020
QY 1021 GATTTTACTACTGTGATGGTGCACAAAGTACATTTTCCCGGATTAAGTACACAGGCA 1080
Db 1021 GATTTTACTACTGTGATGGTGCACAAAGTACATTTTCCCGGATTAAGTACACAGGCA 1080
QY 1081 GGAGCGAANTANTGGAAGAAAGGGAGGCTCCCAACTTTGACAAACAATATCAAG 1140
|||||

Db 1081 GGAGCGAANTANTGGAAGAAAGGGAGGCTCCCAACTTTGACAAACAATATCAAG 1140
QY 1141 GGTCTTTGATTAATCACTTTTGTGATGATTTTCCAAAAGACGTTTACAGAGAGACG 1200
Db 1141 GGTCTTTGATTAATCACTTTTGTGATGATTTTCCAAAAGACGTTTACAGAGAGACG 1200
QY 1201 AGAGAGGATCAAAAGCTCTGAAACAAAGGTCAGGCAAGGATATCAATGAGACTG 1260
Db 1201 AGAGAGGATCAAAAGCTCTGAAACAAAGGTCAGGCAAGGATATCAATGAGACTG 1260
QY 1261 CAAGGATTTGAGAGTGAATTAATTTGACTTTTAAATTAAGTAAAGCATATT 1320
Db 1261 CAAGGATTTGAGAGTGAATTAATTTGACTTTTAAATTAAGTAAAGCATATT 1320
QY 1321 TATTAATCGAAGGTTTTTTGTGTGTTTTTTGTTTTTATTTCAATGCAAGT 1376
Db 1321 TATTAATCGAAGGTTTTTTGTGTGTTTTTTGTTTTTATTTCAATGCAAGT 1376

RESULT 3
US-09-388-993-2
; Sequence 2, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAB-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
; US-09-388-993-2

Query Match 99.9%; Score 1374; DB 5; Length 1376;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTACCGGGAGCTCGGGAGCTCCCGGGAAGTGCACCGGCAAGAGAGGGGGCTAGCTACTG 60
Db 1 TCTACCGGGAGCTCGGGAGCTCCCGGGAAGTGCACCGGCAAGAGAGGGGGCTAGCTACTG 60


```

OY 61 TCTCTGCGGACCAGGAGAGACCCCGCCCGCCGCTGTGAGGCGGCGCCACACAGGGCGG 120
DB 61 TCTCTGCGGACCAGGAGAGACCCCGCCCGCCGCTGTGAGGCGGCGCCACACAGGGCGG 120
OY 121 GTGGGCTGGCGAGCCGAGCGGCGGCGAGAGAGGCTGTGAGAGTGTGTGAACAGAGACC 180
DB 121 GTGGGCTGGCGAGCCGAGCGGCGGCGAGAGAGGCTGTGAGAGTGTGTGAACAGAGACC 180
OY 181 CGGACAGAGAGAACCATGGCTCCGAGAACCGACACCTTTTGGCTGTGTCTATAC 240
DB 181 CGGACAGAGAGAACCATGGCTCCGAGAACCGACACCTTTTGGCTGTGTCTATAC 240
OY 241 CTCAATCGGGGCGGTGATTCGCGAGACGATTTCTATAGATCTTGGGGTCTCTGAGT 300
DB 241 CTCAATCGGGGCGGTGATTCGCGAGACGATTTCTATAGATCTTGGGGTCTCTGAGT 300
OY 301 GCGCTATATAAGATTTAAAAAGCGCTATAGAAAACTAGCCCTGACCTTCATCCGAC 360
DB 301 GCGCTATATAAGATTTAAAAAGCGCTATAGAAAACTAGCCCTGACCTTCATCCGAC 360
OY 361 CGGAACCCCTGATGATCCAGAGCCGAGAGAAATTCAGATCTGGGTGCTTATGAG 420
DB 361 CGGAACCCCTGATGATCCAGAGCCGAGAGAAATTCAGATCTGGGTGCTTATGAG 420
OY 421 GTTCTGATAGTGTAGAGAAACGAAACAGTACATCTTATGCTGAAGAAGATTAAA 480
DB 421 GTTCTGATAGTGTAGAGAAACGAAACAGTACATCTTATGCTGAAGAAGATTAAA 480
OY 481 GATGGTATCAGAGCTCCCATGAGACATTTTTCACACTCTTTGGGGATTTGGTTTC 540
DB 481 GATGGTATCAGAGCTCCCATGAGACATTTTTCACACTCTTTGGGGATTTGGTTTC 540
OY 541 ATGTTTGAGAGAACCCCTGCTCAGCAGACAGAAATATTCACAGAGAGATGATATT 600
DB 541 ATGTTTGAGAGAACCCCTGCTCAGCAGACAGAAATATTCACAGAGAGATGATATT 600
OY 601 GTATATCTAGAGTCACTTTTGGAAAGATATGACAGAAATTTTGTGAAGTATTAGA 660
DB 601 GTATATCTAGAGTCACTTTTGGAAAGATATGACAGAAATTTTGTGAAGTATTAGA 660
OY 661 AACAAACCTGTGCGAAGCGAGGCTCCGGAACCGAGAGGCAATTCGCGCAAGAGATG 720
DB 661 AACAAACCTGTGCGAAGCGAGGCTCCGGAACCGAGAGGCAATTCGCGCAAGAGATG 720
OY 721 CGGACCAACCAAGCTGGGCGCTGCGCAATGACCCAGAGAGTGTCTCGACGAA 780
DB 721 CGGACCAACCAAGCTGGGCGCTGCGCAATGACCCAGAGAGTGTCTCGACGAA 780
OY 781 TGGCCTTAATGTCAAACTAGTGAATGAAGACGAAACGCTGGAATAGAGCTGGG 840
DB 781 TGGCCTTAATGTCAAACTAGTGAATGAAGACGAAACGCTGGAATAGAGCTGGG 840
OY 841 GTGAGAGACGCGATGAGTACCCCTTATGAGAAAGGTAGGCTACGCGATGGGGAG 900
DB 841 GTGAGAGACGCGATGAGTACCCCTTATGAGAAAGGTAGGCTACGCGATGGGGAG 900
OY 901 CCGGAGATTTACGGTCCGAAATCAAAAGTTGTCAGACCCCAATATTTGAAGAGAGA 960
DB 901 CCGGAGATTTACGGTCCGAAATCAAAAGTTGTCAGACCCCAATATTTGAAGAGAGA 960
OY 961 GATGATTTGTACAAATGTGACAGTCTATTAAGTGAATGATCTGTTGGCTTTGAGATG 1020
DB 961 GATGATTTGTACAAATGTGACAGTCTATTAAGTGAATGATCTGTTGGCTTTGAGATG 1020
OY 1021 GATATTACTACTTGTGATGTCACAAAGTACATATTTCCCGGATTAAGTACACAGGCA 1080
DB 1021 GATATTACTACTTGTGATGTCACAAAGTACATATTTCCCGGATTAAGTACACAGGCA 1080
OY 1081 GAGAGCAANTANTGAGAAAGAGGGAGGCTCCCACTTTTGACAAACAATATCAAG 1140
DB 1081 GAGAGCAANTANTGAGAAAGAGGGAGGCTCCCACTTTTGACAAACAATATCAAG 1140

```

```

OY 1141 GGCCTTTGATTAATCACTTTTGAATGAGATTTCCAAAAGACGTTAAACAGAGAACGC 1200
DB 1141 GGCCTTTGATTAATCACTTTTGAATGAGATTTCCAAAAGACGTTAAACAGAGAACGC 1200
OY 1201 AGAAGAGTATCAAAACAGCTCTGAAACAGGGTCACTGAGAGAGTATCAATGAGACTG 1260
DB 1201 AGAAGAGTATCAAAACAGCTCTGAAACAGGGTCACTGAGAGAGTATCAATGAGACTG 1260
OY 1261 CAAGGATTTGAGAGTGAATATAAATTTGACTTTGTTTAAATAAGTAAATACGATATT 1320
DB 1261 CAAGGATTTGAGAGTGAATATAAATTTGACTTTGTTTAAATAAGTAAATACGATATT 1320
OY 1321 TATATCTGCAAGGTTTTTGTGTGTGTTTGTGTTTATTTTCAATATGCAAGT 1376
DB 1321 TATATCTGCAAGGTTTTTGTGTGTGTTTGTGTTTATTTTCAATATGCAAGT 1376

RESULT 4
US-08-868-288A-4
; Sequence 4: Application US/08868288A
; Patent No. 592567
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868,288A
; FILING DATE: June 3, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
; US-08-868-288A-4

Query Match 6.6%; Score 91.4; DB 4; Length 1350;
Best Local Similarity 66.5%; Pred. No. 3.8e-17;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;
OY 268 GATTCTATTAAGATCTTGGGCGCTCGAAGTGCCTATATAAGGATATTAAAGGCC 327
DB 163 GATTACTATGAAGTTCTAGGCGCTGACAGACATGCTTACCCGAGATATTAAAGGCA 222
OY 328 TATAGAAACTAGCCCTGCACTTCATCCGACCGGAACCTG--ATGATCCACAAGCC 384

```

Db 223 TATCGGAACTGGCAGTGGCATCCAGATAAAAATCTGAGATATAAGAAAGCA 282
QY 385 CAGAGAAATTCAGATCTGGGTGCTTATGAGTCTGTGCAGATAGTAAAGCG 444
Db 283 GAGAGAAATTCAGAGATGAGCGGAGCATATGAAAGTCTGTGCGATAGGAAAGCG 342
QY 445 AAACAGTAGCTACTTATGAGTGAAGAGATTAAGATGG 485
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGCTGG 383

RESULT 5

US-09-235-373-4
; Sequence 4, Application US/09235373
; Patent No. 6001396
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAM-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMT2RAT01
; CLONE: 260873
US-09-235-373-4

Query Match 6.6%; Score 91.4; DB 5; Length 1330;
Best Local Similarity 66.5%; Pred. No. 3.8e-17;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATTAAGATCTTGGGGTGCCTCGAAGTGCCTCTATAAGATATTAAGAGCC 327
Db 163 GATTACTATGAAGTTCTAGCGGTGAGAGACATGCCCTCACCGAGATATTAAGAGCA 222
QY 328 TATAGAACTAGCCCTGCGAGCTTCATCCGACCGGAAACCTG---ATGATCCACAAGCC 384
Db 223 TATCGGAACTGGCAGTGAAGTGCATCCAGATTAAGATCTGAGATTAAGAGAGCA 282
QY 385 CAGAGAAATTCAGAGATCTGGGTGCTTATGAGTCTGTGCAGATAGTAAAGAGCG 444
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGCTGG 383

Db 283 GAGAGAAATTCAGAGATGAGCGGAGCATATGAAAGTGTGCGATGCTAAGAAAGCG 342
QY 445 AAACAGTAGCTACTTATGAGTGAAGAGATTAAGATGG 485
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGCTGG 383

RESULT 6

US-09-388-993-4
; Sequence 4, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAM-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMT2RAT01
; CLONE: 260873
US-09-388-993-4

Query Match 6.6%; Score 91.4; DB 5; Length 1330;
Best Local Similarity 66.5%; Pred. No. 3.8e-17;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATTAAGATCTTGGGGTGCCTCGAAGTGCCTCTATAAGATATTAAGAGCC 327
Db 163 GATTACTATGAAGTTCTAGCGGTGAGAGACATGCCCTCACCGAGATATTAAGAGCA 222
QY 328 TATAGAACTAGCCCTGCGAGCTTCATCCGACCGGAAACCTG---ATGATCCACAAGCC 384
Db 223 TATCGGAACTGGCAGTGAAGTGCATCCAGATTAAGATCTGAGATTAAGAGAGCA 282
QY 385 CAGAGAAATTCAGAGATCTGGGTGCTTATGAGTCTGTGCAGATAGTAAAGAGCG 444
Db 283 GAGAGAAATTCAGAGATGAGCGGAGCATATGAAAGTCTGTGCGATAGGAAAGCG 342
QY 445 AAACAGTAGCTACTTATGAGTGAAGAGATTAAGATGG 485
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGCTGG 383

RESULT 7

US-08-974-546-2
; Sequence 2, Application US/08974546
; Patent No. 5945287
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purni
; TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,546
; FILING DATE: Filed Herewith
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0428
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITWY21
; CLONE: 2525691
; US-08-974-546-2

Query Match 4.7%; Score 65.2; DB 4; Length 2349;
Best Local Similarity 56.3%; Pred. No. 2.6e-09;

Matches 143; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

OY 262 GGACGAGATTCTTAAGATCTGGGGGCTCCGAAAGCCCTCTTAAGGATTATAA 321
DB 105 GGAAGAGATTATTAAGATCTGGGATCCCATCGGGGCGCAAGAGATGATCAAG 164
OY 322 AAGGCTATAGAACTACCCCTGACCTTCAATCCCGACCGAACCCTGATGACACA 381
DB 165 AAGGCTACCGGAAGATGCGCTTGAAGTACACCCGACGAAAGAA---TAAAGACCCCAAC 221
OY 382 GCCCAGGAAGAAATTCAGATCTGGGCTGCTTATGAGTTCTGTGACAGTAGAGAA 441
DB 222 GCTGAGGAGAAAGTTTAAAGAGATTGCGAGGCGCTATGATGCTAAGTAGAACCCCAAGAA 281
OY 442 CGAAGACAGTACGATCTATGATGTAAGAGATTAAGAGATGATGATGATGATGAT 501
DB 282 CGGGGCTTATGACACGATGAGGAGAGAGGCTGAAAGACCGGCGGCTGACATCAGGT 341
OY 502 GGAGACATTTTTC 515
DB 342 GGCTCCAGTGGCTC 355

RESULT 8

US-08-486-955A-6
; Sequence 6, Application US/08486955A
; Patent No. 5747299
; GENERAL INFORMATION:
; APPLICANT: FATHMAN, Garrison
; APPLICANT: BLOOM, Debra
; TITLE OF INVENTION: Anergy Genes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,955A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A59741-1
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 672 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-486-955A-6

Query Match 4.7%; Score 65; DB 2; Length 672;
Best Local Similarity 56.9%; Pred. No. 1.3e-09;

Matches 119; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

OY 256 ATTGCCGACGAGATTTCTATTAAGATCTGGGGGCTCCGAAAGCCCTTATAAGGAT 315
DB 31 ACTTCCGGGGAATCGTTATACCATCTTCTTGACCTGACAAAGATGCAACCTCAGATGAC 90
OY 316 ATTAAGAGGCTTATAGAACTACCCCTGACGCTTCAATCCCGACCGAACCCTGATGAT 375
DB 91 ATTAAGAGGCTTATAGAACTACCCCTGACGCTTCAATCCCTGACAAAGACCTGATTAAC 150
OY 376 CCAGAGCCCGAAGAAATTCAGATCTGGGCTGCTGTTATGAGTTCTGTGAGTACT 435
DB 151 CCAGAGGCTGACGACAAAGTTTAAGAGATTAAACAACGACACGCCATCTTGACAGAGCC 210
OY 436 GAGAAACGAAACAGTACGATCTATGAG 464
DB 211 ACAGAAAGAAACATTTATGACAGATATG 239
RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHIEFLINGER, F.
; APPLICANT: FALKNER, F. G.

[illegible]

```

1 ADDRESSSEE: Pennie & Edmonds
2 STREET: 1155 Avenue of the America
3 CITY: New York
4 STATE: New York
5 COUNTRY: USA
6 ZIP: 10036/2711
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Diskette
9 COMPUTER: IBM Compatible
10 OPERATING SYSTEM: DOS
11 SOFTWARE: FASTSED Version 2.0
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/221,816B
14 FILING DATE: 01-APR-1994
15 CLASSIFICATION: 435
16 ATTORNEY/AGENT INFORMATION:
17 NAME: COITZEL, Laura A
18 REGISTRATION NUMBER: 30,742
19 REFERENCE/DOCKET NUMBER: 7960-030
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: (212) 790-9090
22 TELEFAX: (212) 869-8864
23
24 TELEX: 66141 PENNIE
25 INFORMATION FOR SEQ ID NO: 1:
26
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 1515 base pairs
29 TYPE: nucleic acid
30 STRANDEDNESS: single
31 TOPOLOGY: linear
32
33 MOLECULE TYPE: DNA
34
35 OS-08-221-816B-1

```

	Query Match	3.9%	Score 53.4	DB 2	Length 1515	
	Best Local Similarity	58.1%	Pred. No. 5.6e-06			
	Matches 118	Conservative	0	Mismatches 76	Indels 9	Gaps 1
QY	263	GACGAGATTTCTATAGACTCTGGGGGTCCTCGAAGTGCCCTATTAAGATATTAAAA	322			
Db	1175	GACGAGATTTATTAACAAATCTTGGGAGTAAAAAAGAAATGCGCAAAAAGCAAAATCATTTA	1234			
QY	323	AGGCTATATAGAAACTAGCCCTGCAGCTTCATCCCGACCCGAAACCCCTGATGATCCAC---	380			
Db	1235	AAGCATACCGAAATTTAGCACTGCAGTGGCACCCAGACAACTTCCAGAACGGAAGAAAA	1294			
QY	380	-----AAGCCCGAGGAAATTCACAGATCTGGGTGCTTATGAGGTTCGTGCAGATA	433			
Db	1295	AGAAAAAAGCTGAGAAAGAAAGTCATTGACATATGCAAGCTGTCTAAAGAAAGTCCTCTCGATTC	1354			
QY	434	GTGAGAAACGGAAACATGATGCAT	456			
Db	1355	CAGAAATGAGGAAGAAATTTGAT	1377			

RESULT 11

US-08-143-219-26

; Sequence 26, Application US/08143219

; Patent No. 5670330

; GENERAL INFORMATION:

; APPLICANT: Sonenberg, Nahum

; APPLICANT: Katze, Michael G.

; APPLICANT: Roy, Sophie

; APPLICANT: Koromilas, Antoine E.

; TITLE OF INVENTION: Tumor-Cell Assay Method and Kit

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 611 West Sixth Street

; CITY: Los Angeles

; STATE: CA

; COUNTRY: USA

; ZIP: 90017

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; OPERATING SYSTEM: PC-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION NUMBER: US/08/143,219

; FILING DATE: October 25, 1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; PRIOR APPLICATION DATA: including application

; APPLICATION NUMBER: 08/141,244

; FILING DATE: October 22, 1993

; APPLICATION NUMBER: 07/953,681

; FILING DATE: September 29, 1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Douglas E. Olson

; REGISTRATION NUMBER: 22,798

; REFERENCE/DOCKET NUMBER: 204/139

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELETYPE: 67-3510

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1687 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; HYPOHETICAL: NO

; ORIGINAL SOURCE:

; INDIVIDUAL ISOLATE: BOVINE P58 GENE, FIGURE 10A

; US-08-143-219-26

; Query Match

; Best Local Similarity 3.9%; Score 53.4; DB 1; Length 1687;

; Matches 118; Conservative 0; Mismatches 76; Indels 9; Gaps 1;

; Db

; 1290 GACGAGATTCTATAGATCTTGGGGTCCCTGTAAGTCTTAAAGTATTAAAA 322

; 323 AGGCTATATGAACTAGCCCTGAGCTTCATCCGACCGGAACCTGATGCCAC-- 380

; Db

; 1350 AAGCATACCGAAATTAATGACATGACGACCCAGACAACTTCCAGAAAGAA 1409

; Oy

; 380 -----AAGCCAGAGAAATCCAGATCTGGTCTGCTTATAGAGTTCTGTAGATA 433

; Db

; 1410 AGAAAAAGCTGAGAAAGTTCATTGACATAGACAGCTGTAAAGAGTCTCTCGATC 1469

Oy 434 GTGAGAAACGGAACAGTACAT 456

Db 1470 CAGAAATGAGAGAGAGAGTTCAT 1492

RESULT 12

US-08-897-340-4

; Sequence 4, Application US/08897340

; Patent No. 595306

; GENERAL INFORMATION:

; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.

; TITLE OF INVENTION: Weight Control Pathway Genes and Uses

; NUMBER OF SEQUENCES: 36

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD, LLP

; STREET: 28 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/08/897,340

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/715,032

; FILING DATE: 17-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Silverl, Jean M.

; REGISTRATION NUMBER: 39,030

; REFERENCE/DOCKET NUMBER: NMT-005CP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1700 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-897-340-4

; Query Match

; Best Local Similarity 3.5%; Score 48; DB 4; Length 1700;

; Matches 69; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

; Db

; 267 AGATTCTATAGATCTTGGGGTCCCTGTAAGTCTTAAAGTATTAAAAAGGC 326

; Db

; 991 AGATTCTATAGATCTTGGGGTCCCTGTAAGTCTTAAAGTATTAAAAAGGC 1050

; Oy

; 327 CTATGGAAGTACGCTGACGCTTCATCCGACCGGAACCTG 370

; Db

; 1051 TTACCGGAAGGCTTGTGATGACATCCAGATCGGACAGTG 1094

; RESULT 13

US-08-472-534-4

; Sequence 4, Application US/08472534

; Patent No. 5919620

; GENERAL INFORMATION:

; APPLICANT: Hamel, Josee

; APPLICANT: Brodeur, Bernard R

; APPLICANT: Martin, Denis

; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM

; STREPTOCOCCUS PNEUMONIAE

```

; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,534
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr, James F
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: Biovac-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4320 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3265..4320
; OTHER INFORMATION: /product= "NH2-terminal portion of
; OTHER INFORMATION: DNA J"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 682..2502
; OTHER INFORMATION: /product= "Heat-Shock Protein 72"
; FEATURE:
; NAME/KEY: mat-peptide
; LOCATION: 682..2502
; US-08-472-534-4

Query Match          3.5%; Score 47.8; DB 4; Length 4320;
Best Local Similarity 54.3%; Pred. No. 0.00048;
Matches 119; Conservative 0; Mismatches 97; Indels 3; Gaps 1;

QY 268 GATTTCATAGATCTTGGGGGCTCGAGAGCTCTATTAAGATATTAAGAGCC 327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3277 GAATTTTATGATCGTGGGGGATTCACAAAGCTTCGCGACAGAAATCAAAAGGCT 3336
QY 328 TATAGGAAGTACCGCTGCGCTTCATCCGACGGAACCTCATGATCCACAAAGCCAG 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3337 TATCGTAGCTTTCACAAAATATCACCCAGATATCAAC--AAGGAGCCTGGTGTGAG 3393
QY 388 GAGAAATTCAGAGATCTGGGTGCTGTATAGAGTCTGTCAAGATAGTGAGAAACGAAA 447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3394 GACAAATACAGAGAGTTCACAAAGCCTATGAGACTTTGAGTGACGACCAAAACGTGCT 3453
QY 448 CAGTACGATCTATGCTGACGAAGCATTTAAAGATGCT 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3454 GCCTATGACACGATATGTGTCTGACAGGCGCCAAATGCTGCT 3492

RESULT 14
```

```

US-08-879-260-3
; Sequence 3, Application US/08879260
; Patent No. 5935851
; GENERAL INFORMATION:
; APPLICANT: Murthy, Anita E.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: TPR-Containing Genes
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/879,260
; FILING DATE: 19JUN1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/020,204
; FILING DATE: 20JUN1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.4260001/7AG/SRL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1478
; US-08-879-260-3

Query Match          3.3%; Score 44.8; DB 4; Length 1756;
Best Local Similarity 64.4%; Pred. No. 0.002;
Matches 67; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 267 AGATTTCATAGATCTTGGGGGCTCGAGAGCTCTATTAAGATATTAAGAGCC 326
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1136 AGATTACTACAGATTTTGTAGAGAGACAGAAATGCTCTGAGGACGAGATCAAGAAAGC 1195
QY 327 CTATAGGAAGTACCGCTGACGCTTCAATCCGACGGAACCCCTG 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1196 TTATCGGAAGCGGCTTGTATGACACATCCAGATCGGCAATAGTG 1239

RESULT 15
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
```

STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckerts, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start=19
OTHER INFORMATION: /function="gene eryA"
OTHER INFORMATION: /product="eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921

LOCATION: 4471..5847
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function="approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start=10723
OTHER INFORMATION: /function="gene eryA"
OTHER INFORMATION: /product="orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function="approximatr span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modu"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921

```
; OTHER INFORMATION: /function="approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function="approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function="approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
; US-07-642-734C-3
```

```
Query Match 3.0%; Score 40.8; DB 2; Length 20235;
Best Local Similarity 54.7%; Pred. No. 0.14;
Matches 81; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 70 ACCAGGAGAGACCCCGCGCCCGCGGTGTGAGGCGGCTTCACAGGGCCGCGTGGCTGG 129
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10933 ATGCGGAGGAGCCCGCGCGCGCGAGCGGCTGGAGCGGAGCCCGCGCGCTGGCGGG 10992
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 130 CGAGCCGACGCGGCGCGCGAGAGGCTGTGAGAGTGTGTGACAGAGACCCGCGGACAGA 189
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10993 ATGCTCGCGGCGCGCGGCGGCGACTTCGACGCGGCGCTTCTCGGATCTGCGCGCGAGGGC 11052
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 190 GGAACCATGGCTCCGCGAGAACTGTAGCA 217
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11053 CTGGCGATGAGCCCGCGAGCAGCGGATCA 11080
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: June 30, 2000, 13:32:40
Job time: 5921 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 11:55:33 ; Search time 70.1 Seconds

(Without alignments)
4911.046 Million cell updates/sec

Title: US-09-501-714-2

Perfect score: 1376
Sequence: 1 TCTACCGGAGCTCGGAGCT.....TTTATTTCATATGCAAGT 1376

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1374	99.9	1376	1 X06099	Human DnaJ-like pr
2	91.4	6.6	1330	1 X06100	Human DnaJ-like pr
3	70	5.1	1204	1 X14194	H. pylori GHP0 885
4	67.2	4.9	28295	1 X20507	Polynucleotide seq
5	65	4.7	672	1 T11788	Mouse cysteine str
6	65	4.7	672	1 T11789	Human cysteine str
7	62.6	4.5	1191	1 Q57429	DNA J like protein
8	60.6	4.4	110000	1 X20248_03	Continuation (4 of
9	59.4	4.3	110000	1 X20248_02	Continuation (3 of
10	59.2	4.3	1726	1 X07570	Homo sapiens fetal
11	57	4.1	1331	1 V86438	EST clone AM610. N
12	53.4	3.9	1515	1 Q77721	p58 CDNA clone. SC
13	53.4	3.9	1515	1 V22052	Bovine p58 DNA. SC
14	53.4	3.9	1687	1 V01072	Bovine p58 gene. S
15	49.4	3.6	6336	1 W74530	Staphylococcus aur
16	48	3.5	1700	1 V11858	Mus musculus Tsp I
17	47.8	3.5	4320	1 T73388	DNA encoding HSP72
18	47.8	3.5	7174	1 V52322	Streptococcus pneu
19	47.8	3.5	110000	1 T58840_0	Mycoplasma genital
20	47.8	3.5	110000	1 T58840_2	Continuation (3 of
21	47.2	3.4	459	1 W87477	EST clone BN379. N
22	46.8	3.4	450	1 Q03633	Mycoplasma hyopneu
23	46.4	3.4	1701	1 X39655	Renal cancer assoc
24	46	3.3	828	1 X30633	H. pylori cytoplas
25	46	3.3	837	1 X30634	H. pylori cytoplas
26	46	3.3	990	1 X14165	H. pylori GHP0 542
27	45	3.3	2717	1 X13034	Enterococcus faeca
28	43.6	3.2	491	1 V02136	Human secreted pro
29	43.6	3.2	491	1 T88058	Partial CDNA clone
30	43.6	3.2	8169	1 V26609	Actinomyadura hdbis
31	42	3.1	201	1 N70194	Signal portion of
32	41.8	3.0	336	1 X40691	Human secreted pro
33	41.6	3.0	201	1 N70195	Streptomycetes prote
34	40.8	3.0	2312	1 X16152	Chicken Sox1 cDNA.

35	40.8	3.0	29879	1 Q46806	eryA region of S.
36	39.6	2.9	3748	1 T84330	Maize DnaJ-related
37	39.6	2.9	3748	1 V72905	Maize DnaJ clone D
38	39.2	2.8	832	1 X04554	DNA encoding a hum
39	38.8	2.8	1860	1 Q50632	Brain factor-2. Is
40	38.6	2.8	1055	1 T15275	E.coli K12 dnaJ co
41	38.6	2.8	5852	1 Q11710	Dictyostelium plas
42	38.4	2.8	284	1 V90426	EST clone DM360. N
43	38	2.8	2277	1 V05370	Human telomerase p
44	38	2.8	2277	1 V13834	Homo sapiens ambig
45	37.8	2.7	997	1 Q48024	ATGF gene #1. Cell

ALIGNMENTS

RESULT 1	
X06099	X06099 standard; DNA; 1376 BP.
ID	X06099: 196..1272
AC	09-APR-1999 (first entry)
DE	Human DnaJ-like protein, HSPJ1 encoding DNA.
KE	DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;
KW	immune disorder; inflammation; tissue damage; diabetes; wound healing;
OS	chromosome mapping; ss.
OS	Homo sapiens.
FT	Key
FT	Location/Qualifiers
FT	196..1272
FT	/*tag= a
FT	/product= "DnaJ-like protein, HSPJ1"
FT	/transl_except= (pos:1087..1089; aa:Xaa)
FT	/transl_except= (pos:1090..1092; aa:Xaa)
FT	/note= "Xaa = unknown"
PD	10-DEC-1998.
PE	02-JUN-1998; U11182.
PR	03-JUN-1997; US-868288.
PA	(INCY-) INCYTE PHARM INC.
PI	Au-Young J, Bandman O, Lal P;
DR	WPI: 99-070259/06.
DR	P-PSDB: W94065.
PT	New nucleic acid encoding human DnaJ-like proteins - for diagnosis,
PT	treatment and prevention of cancer, immune disorders and
PT	inflammation
PS	Claim 6; Fig 1A-D; 73pp; English.
CC	This DNA encodes a human DnaJ-like protein, HSPJ1. The invention provides
CC	two human DnaJ-like proteins which are heat shock proteins J1 and J2
CC	(HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host
CC	cells containing a vector comprising the nucleic acids are used for the
CC	production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used
CC	to raise Ab, therapeutically and to screen for specific binding agents.
CC	Antagonists are used to treat or prevent a wide variety of solid cancers,
CC	leukaemia and lymphoma; immune disorders (typical of many disclosed are
CC	acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
CC	diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
CC	and parasitic infections) and inflammation. Agonists may be used to treat
CC	or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
CC	attack, ischemia, UV exposure, insulin-dependent diabetes and for wound
CC	healing, and may be expressed from a vector containing the nucleic acids.
CC	Fragments of the nucleic acids are used as primers and probes for
CC	detecting and quantifying the HSPJ-encoding nucleic acid in usual
CC	hybridisation and/or amplification assays, therapeutically as antisense,
CC	triplex-forming or ribozyme molecules, and for chromosome mapping.
CC	Sequence 1376 BP; 387 A; 268 C; 397 G; 322 T;
SO	
Query Match	99.9%; Score 1374; DB 1; Length 1376;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 TCTACCGGAGCTCGGAGCTCGGAGCTCGGAGCTCGGAGCTAGCTAGCTG 60
DB	1 TCTACCGGAGCTCGGAGCTCGGAGCTCGGAGCTCGGAGCTCGGAGCTAGCTAGCTG 60

QY 61 TCTCTGCGGACGAGGAGACCCCGCGCCCGGTGAGGGGCGCTCACAGGCGCG 120
 DB 61 TCTCTGCGGACGAGGAGACCCCGCGCCCGGTGAGGGGCGCTCACAGGCGCG 120
 QY 121 GTGGGCTGCGAGCGCGCGCGGAGAGAGGCTGTGAGAGTGTGTGACAGCAGC 180
 DB 121 GTGGGCTGCGAGCGCGCGCGGAGAGAGGCTGTGAGAGTGTGTGACAGCAGC 180
 QY 181 CGGGACAGAGAACCATGGCTCCGAGAACTGAGACCTTTTGGCTTCTGCTATAC 240
 DB 181 CGGGACAGAGAACCATGGCTCCGAGAACTGAGACCTTTTGGCTTCTGCTATAC 240
 QY 241 CTCAATCGGGGCGGTGATTCGCGAGAGATTTCTATAAGATCTGGGGGCTCGTGAAGT 300
 DB 241 CTCAATCGGGGCGGTGATTCGCGAGAGATTTCTATAAGATCTGGGGGCTCGTGAAGT 300
 QY 301 GCCCTCTATAAGATTTAAAAAGGCTATAGAAACTAGCCCTGCAAGCTTCAATCCGAC 360
 DB 301 GCCCTCTATAAGATTTAAAAAGGCTATAGAAACTAGCCCTGCAAGCTTCAATCCGAC 360
 QY 361 CGGAACCCCTGATATCCAGCCAGAGAAATTCAGATCTGGGCTGCTTATGAG 420
 DB 361 CGGAACCCCTGATATCCAGCCAGAGAAATTCAGATCTGGGCTGCTTATGAG 420
 QY 421 GTTCTGTCAGATAGTGAAGAAACGAAACAGTACGATCTTATGCTGAAGAGATTA 480
 DB 421 GTTCTGTCAGATAGTGAAGAAACGAAACAGTACGATCTTATGCTGAAGAGATTA 480
 QY 481 GATGTCATGAGATCCCATGAGACATTTTTCACACTTCTTGGGATTTGGTTTC 540
 DB 481 GATGTCATGAGATCCCATGAGACATTTTTCACACTTCTTGGGATTTGGTTTC 540
 QY 541 ATGTTGGAGAACCCCTCGTCAGCAGACAGAAATATTCACAGAGAGATATATT 600
 DB 541 ATGTTGGAGAACCCCTCGTCAGCAGACAGAAATATTCACAGAGAGATATATT 600
 QY 601 GTATATCTAGAAATGCTATTTGGAAAGAAATATTCAGAGAAATTTTGGAAAGTATGA 660
 DB 601 GTATATCTAGAAATGCTATTTGGAAAGAAATATTCAGAGAAATTTTGGAAAGTATGA 660
 QY 661 AACAAACCTGCGGCAAGGCGCTCCGCAAGCGAAGTGCATTTCCGCAAGAGATG 720
 DB 661 AACAAACCTGCGGCAAGGCGCTCCGCAAGCGAAGTGCATTTCCGCAAGAGATG 720
 QY 721 CGGACACCCAGCTGGGCGCTTCCAAATGACCCAGAGAGTGTCTCGAGAA 780
 DB 721 CGGACACCCAGCTGGGCGCTTCCAAATGACCCAGAGAGTGTCTCGAGAA 780
 QY 781 TGGCCTTATATCTCAAACTAGTGAAGAAAGCAAGCTGGAAGTGAATAGACCTGGG 840
 DB 781 TGGCCTTATATCTCAAACTAGTGAAGAAAGCAAGCTGGAAGTGAATAGACCTGGG 840
 QY 841 GTGAGAGAGCGATGAGTACCCCTTATTTGGAAGAGTACCTCACCGATGGGGAG 900
 DB 841 GTGAGAGAGCGATGAGTACCCCTTATTTGGAAGAGTACCTCACCGATGGGGAG 900
 QY 901 CCTGAGAGATTTAGCGTTCGCAATCAAAAGTGTGCAACACCCCAATATTTGAAGAGAGA 960
 DB 901 CCTGAGAGATTTAGCGTTCGCAATCAAAAGTGTGCAACACCCCAATATTTGAAGAGAGA 960
 QY 961 GATGATTTGACACAAATGTGAGAGTCTATTTAGTTGAGTCACTGGTTGGATGAGT 1020
 DB 961 GATGATTTGACACAAATGTGAGAGTCTATTTAGTTGAGTCACTGGTTGGATGAGT 1020
 QY 1021 GATATTACTACTTGGATGTCACAGTACATATTTCGCGGATTAAGATCCAGGCA 1080
 DB 1021 GATATTACTACTTGGATGTCACAGTACATATTTCGCGGATTAAGATCCAGGCA 1080
 QY 1081 GGAAGCAANTANTAGGAAGAGGAGGCTCCCACTTTGACAAACAATATCAAG 1140
 DB 1081 GGAAGCAANTANTAGGAAGAGGAGGCTCCCACTTTGACAAACAATATCAAG 1140

QY 1141 GCCTCTTGTATATCACTTTTGTATGTCATTTTCCAAAGACAGTTTACAGAGAGCG 1200
 DB 1141 GCCTCTTGTATATCACTTTTGTATGTCATTTTCCAAAGACAGTTTACAGAGAGCG 1200
 QY 1201 AGAAGAGTATCAAAACGCTACTGAACAGGGGCTGTCAGAGAGTATCAATGACAGT 1260
 DB 1201 AGAAGAGTATCAAAACGCTACTGAACAGGGGCTGTCAGAGAGTATCAATGACAGT 1260
 QY 1261 CAAGGATATTTGAGAGTAAATTTGACCTTTGTTTAAATATGTAATTAAGCATATTT 1320
 DB 1261 CAAGGATATTTGAGAGTAAATTTGACCTTTGTTTAAATATGTAATTAAGCATATTT 1320
 QY 1321 TATTATCTGCAAGGTTTTTTGTGTGTGTTTGTATTTTCAATATGCAAGT 1376
 DB 1321 TATTATCTGCAAGGTTTTTTGTGTGTGTTTGTATTTTCAATATGCAAGT 1376

RESULT 2
 ID X06100
 AC X06100 standard: DNA; 1330 BP.
 DT 09-APR-1999 (first entry)
 DE Human dnaJ-like protein, HSPJ2 encoding DNA.
 KW dnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;
 KW immune disorder; inflammation; tissue damage; diabetes; wound healing;
 OS Homo sapiens.
 FH Key
 FT CDS
 FT 157..1149
 FT /tag= a
 FT /product= "dnaJ-like protein, HSPJ2"
 FT /transl_except- (pos:871..873; aa:Xaa)
 FT /transl_except- (pos:970..972; aa:Xaa)
 FT /transl_except- (pos:1024..1026; aa:Xaa)
 FT /transl_except- (pos:1027..1029; aa:Xaa)
 FT /transl_except- (pos:1033..1035; aa:Xaa)
 FT /transl_except- (pos:1048..1050; aa:Xaa)
 FT /transl_except- (pos:1057..1059; aa:Xaa)
 FT /transl_except- (pos:1060..1062; aa:Xaa)
 FT /transl_except- (pos:1069..1071; aa:Xaa)
 FT /transl_except- (pos:1081..1083; aa:Xaa)
 FT /note= "Xaa = unknown"
 FT PN W0985509-A2.
 PD 10-DEC-1998.
 PF 02-JUN-1998: U11182.
 PR 03-JUN-1997: US-868288.
 PA (INCYTE-) INCYTE PHARM INC.
 PI Au-Young J, Bandman O, Lal P;
 DR WPI; 99-070259/06.
 DR P-PSDB; W94066.
 PT New nucleic acid encoding human dnaJ-like proteins - for diagnosis,
 PT treatment and prevention of cancer, immune disorders and
 PT inflammation
 PS Claim 27; Fig 3A-D; 73pp: English.
 CC This DNA encodes a human dnaJ-like protein, HSPJ2. The invention provides
 CC two human dnaJ-like proteins which are heat shock proteins J1 and J2
 CC (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host
 CC cells containing a vector comprising the nucleic acids are used for the
 CC production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used
 CC to raise Ab, therapeutically and to screen for specific binding agents.
 CC Antagonists are used to treat or prevent a wide variety of solid cancers,
 CC leukaemia and lymphoma; immune disorders (typical of many diseases are
 CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
 CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
 CC and parasitic infections) and inflammation. Antagonists may be used to treat
 CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
 CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound
 CC healing, and may be expressed from a vector containing the nucleic acids.
 CC Fragments of the nucleic acids are used as primers and probes for
 CC detecting and quantifying the HSPJ-encoding nucleic acid in usual
 CC hybridisation and/or amplification assays, therapeutically as antisense,
 CC triplex-forming or ribozyme molecules, and for chromosome mapping.
 CC Sequence 1330 BP; 336 A; 318 C; 271 T;

T11788

RESULT	6
ID	T11789
AC	T11789 standard; cDNA; 672 BP.
DT	07-APR-1996 (first entry)
DE	Human cysteine string protein gene.
KW	Human; cysteine string protein; anergy; T-lymphocyte;
KW	differential display; cDNA; reverse transcription; probe;
KW	polymerase chain reaction; cloning; gel electrophoresis;
KW	antibody; diagnostic; immunocassay; autoimmune disease;

RESULT	7
057429	
ID	057429 standard; cDNA to mRNA; 1191 BP.
AC	Q57429; standard; cDNA to mRNA; 1191 BP.
DT	19-OCT-1994 (first entry)
DE	DNA y like protein.
KW	Human cDNA; library; enzyme; protein; ss.
OS	Homo sapiens.
PN	WO9403599-A.
PD	17-FEB-1994.
PF	04-AUG-1993; J01095.
PR	04-AUG-1992; JP-208077.
PR	13-NOV-1992; JP-327619.
PR	26-FEB-1993; JP-061431.
PA	(SAGA) SAGAMI CHEM RES CENTRE.
PI	Iwahori A, Kato S, Kato T, Kim N, Oh S, Sekine S;
DR	WPI: 94-065688/08.
DR	P-PSDB; R46090.
PT	cDNA of human origin and proteins coded by it - which may be

5Q Sequence 1191 BP; 397 A; 207 C; 309 G; 278 T;

Qy	450	GTACGATCTATTG	GTGAAGAAGAT	TAAAGATG	4866
Db	189	ATATGACAAAGAG	GAGAGAACAGG	CAATTAAAGAG	225

WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Oy 267 AGATTTCTAAGAATCTTGGGGGTGCGAGTGCCCTCATATAAGGATATTAAAAAGC 326
||||| |||| | ||||| | ||||| | ||||| |
Db 81529 AGATTATTATGAATTTTTGGGGCTCTCAAAAAGGACCTCAAAAGATAGATAAAAAAGC 81588

QY	327	CFATAGGAACTAGCCCTGCACGCTTCATCCGACCGGAACCTGATGATCCACAAGCCCA	386
Db	81589	TTATAGAAAATATAGCAATTAAATATACACCCAGACAGAAATCAAGGGAATGAAGAAGCCGC	81648

QY 387 GGAGAAATTCAGAGACTGGTGCTGTTATGAGGTTCTGCAGATAGTAGAAGCGAA 446
- - - - -
Db 81649 CTCATCTTTAAAGAAAGCCACACTCAGGCTTACGAAATTTTATAGATGCAATATAAAAAAGC 81708

QY	447	ACAGTACGATCTATTAGTGTGAGAGAGATTAAACATGG	485
		- - - - -	
Do	81709	TAAATACGACGATTGGGCGATTCGCTTTTGAAGGAGG	81747

RESULT 9
X20248-02
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide :
MP Sequence split into 10 fragments LOCUS X20248 Accession X20248

Query Match	4.3%	Score 59.4	DB 1	Length 110000
Best Local Similarity	51.3%	Pred. No. 7.5e+06		
Matches 138, Conservative	0	Mismatches 131,	Indels 0	Gaps 0

Oy 267 AGATTCTAATAAGATCTTGGGSGTGCCTCGAAGTGCCTCTAATAAAGATATTAAAAAGC 326

Db 15206 AGACTACTACATATCTTGTGATACAATAAAAGCTAGTAATGAGGAAATAAAAAAGC 1522

Oy 327 CTATAGGAAACTAGCCCTGCACGTTCANTCCGACCGGAAOCTGANCATCACAAGCCCA 386
 | | | | | | | | | | | | | | |
Db 1526 TTACAAAAAATTGGCAATAAAAATTATCACCCAGACAAAAACAAGGAACCAAATAGCTGA 15325

[illegible]

QY 447 ACAGTAGCACTTATGGTGAAGNAGATTAAAGATGTCATCAGAGCTCCATGGAGA 506
 | | | | | | | | | | | | | | | | | |
Db 15386 AAATATGACTCTTGCGTAACCAAAATTTTAATGGCAACAGACCATTTTGAAAGAGA 15445

QY 507 CATTTTTCACACTTCTTTGGGGATTTTG 535
 || ||||| |||||
 Db 15446 ATTTAGCAGCACACAGATTTGGCAATTTTG 15474

RESULT	10
X07570	
ID	X07570 standard; cDNA; 1726 BP.
AC	X07570;
DT	21-JUN-1999 (first entry)
DE	Homo sapiens fetal kidney clone AM610 secreted protein gene.
KW	Secreted protein; fetal kidney; ds.
OS	Homo sapiens.
FX	
FT	Key
FT	Location/Qualifiers
FT	30..701

FT WO9900405-A1. /product- "secreted protein"
PN 07-JAN-1999.
PD 29-JUN-1998; UI3530.
PF 30-JUN-1997; US-885610.
PR (GEMY) GENETICS INST INC.
PA Agostino MJ, Evans C, Jacobs K, Lavallie ER, McCoy JM,
PI Merberg D, Racie LA, Treacy M;
PT WPI: 99-095671/08.
DR P-PSDB: W95710.
PT New polynucleotides encoding secreted human proteins - are derived
PT from foetal kidney or adult retina cDNA libraries, used as, e.g.
PI potential vaccines
PS Disclosure: Pages 55-56: 76pp: English.
CC The sequence is that encoding a secreted protein from a human fetal
CC kidney clone AM610. Such a sequence is predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. It is also stated to be useful for gene
CC therapy.
S0 Sequence 1726 BP; 543 A; 285 C; 315 G; 583 T;

Query Match 4.3%; Score 59.2; DB 1; Length 1726;
Best Local Similarity 51.2%; Pred. No. 9.9e-07;
Matches 164; Conservative 0; Mismatches 153; Indels 3; Gaps 1;

QY 169 TGAAGACAGACCCGAGAGAACCATGGCTCCGAGAACCGAGACCTTTGGCTG 228
DB 6 TGGACACGAGGCGAGATATTAGAAATGGCTACTCCCGCATCAATTTTCATCTTGCATC 65
QY 229 TTGCTGTATACCTCATCGGGCGGTGATGCCGAGAGATTTCTATTAAGATCTTGGGG 288
DB 66 TGCATTTTAAATGATACAGATTAATTTCTGCTCCCAAAAGCTACTATGATATCTTAGT 125
QY 289 GTGGCTGGAAGTGGCTCTATTAAGATTTTAAAGAGCTTATAGAAAGCTTGGCTGCG 348
DB 126 GTGCAAAATCGGATCAGACCGCCAAATCAAGAGGCTTTTCAAGTTGGCATGAG 185
QY 349 CTTATATCCGACCGAACCTGATGATCAACAGCCAGAGAAATTCAGAGATCTGGT 408
DB 186 TACACACCTGACAAAA---TAAAGCCAGATCTGAGCAAAATTCAGAGATTTGCA 242
QY 409 GCTGCTTATGAGTCTGTGATGATGAGAAAGCAAGATGATGATGATGATGATGATG 468
DB 243 GAAGCATATGAAACATCTCAGATGATGATGATGATGATGATGATGATGATGATGATG 302
QY 469 GAAGGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 488
DB 303 AGTCTTTTACTAGTGTAA 322

RESULT 11

VS6438
ID V86438 standard; cDNA; 331 BP.
AC V86438;
DE 27-APR-1999 (first entry)
KW EST clone AM610.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemokines; chemokines; haemostasis; gene therapy; thrombolytics;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN W09845435-A2.
PD 15-OCT-1998.
PR 10-APR-1998; 006954.
PR 10-APR-1997; US-835913.
PA (GENE) GENETICS INST INC.
PI Agostino M, Jacobs K, Lavallee ER, McCoy JM, Merberg D,
PI Ragle LM, Spaulding V, Treacy M;
PI WPI: 99-070076/06.
DR New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1, Page 234, 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemokine/chemokine activity, haemostatic
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
S0 Sequence 331 BP; 115 A; 69 C; 68 G; 79 T;

Query Match 4.1%; Score 57; DB 1; Length 331;
Best Local Similarity 52.2%; Pred. No. 1.7e-06;
Matches 151; Conservative 0; Mismatches 135; Indels 3; Gaps 1;

QY 200 CTCGAGAACCTGAGACCTTTGGCTGCTGCTGCTATACCTCATCGGGCGGTGATTTG 259
DB 36 CTCGCCAGTCAATTTTCATCTTGCAGATGCAATTTTAAATGATACAGATTAATCTCG 95
QY 260 CCGAGACAGATTTCTATTAAGATCTTGGGGTGGCTCGAGATGCTCTATTAAGATATTA 319
DB 96 CTTCAAAAGCTACTATGATATGATGATGATGATGATGATGATGATGATGATGATGATG 155
QY 320 AAAAGGCTTATAGAAAGTACCTGAGCTTATATCCAGCGGAACCTGATGATGATGATG 379
DB 156 AGAAGGCTTTTCAAAAGTGGCCATGAGACACCTGACAAAA---TAAAGGCCAG 212
QY 380 AACCCAGAGAAATTCAGATCTGGTGGCTGCTTATGAGTGTCTGATGATGATGATGATG 439
DB 213 ATGCTGAGCAAAATTCAGAGATTTGAGAGATTTGAGAGATTTGAGAGATTTGAGATG 272
QY 440 AACGGAACAGTACATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 488
DB 273 GACGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 321

RESULT 12

ID 077721 standard; cDNA; 1515 BP.
AC 077721;
DE 05-JUL-1995 (first entry)
DE p58 cDNA clone.
KW Influenza virus; p58; probe; MDBK cell line; RACE-PCR; ss.
OS Influenza virus.
PN W09423041-A.
PD 13-OCT-1994.
PR 01-APR-1994; 003623.
PR 02-APR-1993; US-042024.
PA (RIBO-) RIBOGENE INC.
PI Katze MG, Mathews MB, Miles VJ, Watson JC, Witherell G;
PI WPI: 94-333201/41.
PT Screening for antiviral agents which have little effect on
PT non-infected cells - by determining whether the agent interacts
PT with a virus or cellular component and allows or prevents
PT preferential translation of viral RNA.
PS Example 6; Page 168-69; 195pp; English.
CC This sequence represents the Influenza virus p58 cDNA. This sequence
CC was isolated using probes based on tryptic peptides derived from p58.
CC These probes were used to screen a cDNA library from MDBK cell line
CC made in lambda Zap II vector. The initial clone contained a long
CC open reading frame but no termination codon suggesting that the 3'
CC end was missing. The 3' end was isolated by RACE-PCR. This allowed
CC construction of the full length p58 cDNA containing 1680 bp.
S0 Sequence 1515 BP; 509 A; 285 C; 362 G; 359 T;

Query Match 3.9%; Score 53.4; DB 1; Length 1515;
Best Local Similarity 58.1%; Pred. No. 3.7e-05;
Matches 118; Conservative 0; Mismatches 76; Indels 9; Gaps 1;

QY 263 GACGAGATTTCTATTAAGATCTTGGGGTGGCTCGAAGTGCCTTATTAAGATATTTAAA 322
DB 1175 GACGAGATTTATTAACAATCTTGGAGATTAAGAAATGCAAAAGCAAGAAATCAATTA 1234
QY 323 AGGCTATAGAAATTAACCTCGAGCTTCATCCGACCGAACCCTGATATATCCAC--- 380
DB 1235 AAGCATACCGAAATTAACCTCGAGCTTCATCCGACCGAACCCTGATATATCCAC--- 1294
QY 380 -----AAGCCAGAGAAATTCAGATCTGGGGTGGCTTATGAGGTTGTCAGATA 433
DB 1295 AGAAAGAGCTGAGAGAGATTCATGACATGACAGCTGCTTAAGAAAGTCTCTCCGATC 1354

Query Match	3.9%	Score 53.4	DB 1	Length 1515
Best Local Similarity	58.1%	Pred. No. 3.7e-05		
Matches 118	Conservative	0	Mismatches 76	Indels 9
				Gaps 1
QY 263	GACGAGATTTTCATTAAGATCTGGGGGGTCCCTCGAAGTGCCCTATTAAGATATTAAAA	322		
Db 1175	GACGAGATTTATATACAAATCTTGGGAGTTAAAAAGAAATGCCCCAAAAGCAATATCATTA	1234		
QY 323	AGGCCTATAGGAAATATACCCCTGCAGCTTCATCCCGACCGGAAACCTGTATCCAC---	380		
Db 1235	AAGCATACCGAAAATTTAGCATCTGCAGTGCCACCCAGACAACTTCCAGAAAGAAAGAAA	1294		
QY 380	-----AAGCCCGAGGAAATTCAGATCTGGGTGCTGCTTATGAGTTCTGTCAGATA	433		
Db 1295	AGAAAAAAGCTGAGAGAAGTTCATTGACATATGACAGCTGCTATAAAGAATGCCCTCGAATC	1354		
QY 434	GTGAGAAACGGAACATGTCAT 456			
Db 1355	CAGAAATGAGGAAGATTTGAT 1377			

RESULT 14

V01072

1D V01072 standard; DNA; 1687 bp.

AC V01072;

DT 30-MAR-1998 (first entry)

DE Bovine P58 gene.

KW Human: PKR; double stranded RNA-activated protein kinase; neoplasms; cell growth; differentiation; tumour suppressor; tumorigenesis; ds.

OS Bos taurus.

Key

Key Location/Qualifiers

FT CDS 116..1630

FT /tag= a

FT /product= P58.protein

FT /transl_except= (pos: 365..367, aa: His)

FT polyA_signal 1626..1631

FT /tag= b

FT /note= "sub-optimal polyA signal"

PN US5670330-A.

PD 23-SEP-1997.

PF 25-OCT-1993; 143219.

PR 25-OCT-1993; US-143219.

PR 29-SEP-1993; US-053681.

PR 22-OCT-1993; US-141244.

PA (UYMC-) UNIV MCGILL.

Query Match	3.9%;	Score 53.4;	DB 1;	Length 1687;
Best Local Similarity	58.1%;	Pred. No. 3.9e-05;		
Matches 118; Conservative	0;	Mismatches 76;	Indels 9;	Gaps 1

RESULT	15
V74530/c	
ID	V74530 standard; DNA; 6336 BP

RESULT	15
ID	V74530/c
AC	V74530 standard; DNA; 6336 BP.
DT	16-MAR-1999 (first entry)
DE	Staphylococcus aureus contig SEQ ID #219.
KW	Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW	skin infection; surgical wound infection; scalded skin syndrome;
KW	toxic shock syndrome; ds.
OS	Staphylococcus aureus.
FH	Key location/Qualifiers
FT	misc_feature
FT	1621..1680
FT	/tag=a
FT	/note=
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	3421..3480
FT	/tag=b
FT	/note=
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	5221..5280
FT	/tag=c
FT	/note=
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	misc_feature
FT	5221..5280
FT	/tag=c
FT	/note=
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:57:47 ; Search time 16.52 Seconds

(without alignments)
288.398 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716
Sequence: 1 MVDYEVGLGVQRHNSPEDIK.....EAERGVEEHVDQROSLDRT 330

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/PCUS.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1696	98.8	330	2	US-08-868-288A-3
2	1696	98.8	330	3	US-09-235-373-3
3	1696	98.8	330	3	US-09-388-993-3
4	635.5	37.0	277	2	US-08-868-288A-7
5	635.5	37.0	277	3	US-09-235-373-7
6	635.5	37.0	277	3	US-09-388-993-7
7	635.5	37.0	277	3	US-08-868-288A-6
8	635.5	37.0	351	3	US-09-235-373-6
9	635.5	37.0	351	3	US-09-388-993-6
10	364.5	21.2	340	2	US-08-974-546-5
11	337	19.6	348	2	US-08-974-546-1
12	311	18.1	397	2	US-08-868-288A-5
13	311	18.1	397	3	US-09-235-373-5
14	311	18.1	397	3	US-09-388-993-5
15	274.5	16.0	419	2	US-08-686-417-3
16	259.5	15.1	352	2	US-08-472-534-6
17	257	15.0	358	2	US-08-868-288A-1
18	257	15.0	358	3	US-09-235-373-1
19	257	15.0	358	3	US-09-388-993-1
20	241.5	14.1	484	2	US-08-879-260-4
21	232.5	13.5	438	2	US-08-897-340-34
22	178	10.4	51	1	US-08-346-849-12
23	178	10.4	51	1	US-08-293-284A-12
24	171.5	10.0	52	1	US-08-346-849-10
25	171.5	10.0	52	2	US-08-293-284A-10
26	160.5	9.4	52	1	US-08-346-849-8
27	160.5	9.4	52	2	US-08-293-284A-8
28	151.5	8.8	52	1	US-08-346-849-9
29	151.5	8.8	52	2	US-08-293-284A-9

30	147	8.6	49	1	US-08-346-849-11	Sequence 11, Appl
31	147	8.6	49	2	US-08-293-284A-11	Sequence 11, Appl
32	110.5	6.4	56	2	US-08-346-849-28	Sequence 28, Appl
33	110.5	6.4	56	2	US-08-293-284A-28	Sequence 28, Appl
34	105	6.1	51	1	US-08-346-849-14	Sequence 14, Appl
35	105	6.1	51	2	US-08-293-284A-14	Sequence 14, Appl
36	103.5	6.0	341	2	US-08-538-711A-8	Sequence 8, Appl
37	103.5	6.0	353	2	US-08-538-711A-7	Sequence 7, Appl
38	103.5	6.0	433	1	US-08-346-849-2	Sequence 2, Appl
39	103.5	6.0	433	2	US-08-293-284A-2	Sequence 2, Appl
40	103.5	6.0	714	2	US-08-990-114-3	Sequence 3, Appl
41	98	5.7	223	1	US-07-667-276A-10	Sequence 10, Appl
42	96.5	5.6	643	2	US-08-551-356-4	Sequence 4, Appl
43	96.5	5.6	643	1	PCT-US93-12687-4	Sequence 4, Appl
44	96	5.6	51	4	US-08-346-849-15	Sequence 15, Appl
45	96	5.6	51	2	US-08-293-284A-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-868-288A-3
: Sequence 3, Application US/08868288A
: Patent No. 5922567
: GENERAL INFORMATION:
: APPLICANT: Au-Young, Janice
: APPLICANT: Lal, Preeti
: APPLICANT: Bandman, Olga
: TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESS: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/868, 288A
: FILING DATE: June 3, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0309 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 330 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: HNT2RAT01
: CLONE: 260873
: US-08-868-288A-3

Query Match 98.8%; Score 1696; DB 2; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.9e-167;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDYEVGLGVQRHNSPEDIKKAYRKALKWHPDKPEKFAERKQVAEYVLSDAK 60

```
|||||
Db 1 MVDYEVLVGVRHASPEDIKAYRKIALKWHDPKNPENKEAEARKKQVAEALEYLSDAK 60
QY 61 KRDIYDKYKGEGLNGGGGSHFDSPEFEFTFRNPDVFRFEGGRDPFSDFEDDPE 120
Db 61 KRDIYDKYKGEGLNGGGGSHFDSPEFEFTFRNPDVFRFEGGRDPFSDFEDDPE 120
QY 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
Db 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
QY 181 GSGMGNFKSIISTYKMNKRIITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
Db 181 GSGMGNFKSIISTYKMNKRIITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
QY 241 ERMRRGQNVLPAPAGLRPPKPPRPSLIRHXPHCLSKKEEGODRPMAPXXMPLASXAG 300
Db 241 ERMRRGQNVLPAPAGLRPPKPPRPSLIRHXPHCLSKKEEGODRPMAPXXMPLASXAG 300
QY 301 XXEGXKRMXAERGVVEEVDQROSLDRT 330
Db 301 XXEGXKRMXAERGVVEEVDQROSLDRT 330
```

```
RESULT 2
US-09-235-373-3
; Sequence 3, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAM-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/668,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-09-235-373-3
```

Query Match 98.8%; Score 1696; DB 3; Length 330;
Best Local Similarity 100.0%; Pred. No. 2,9e-167;

```
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDYEVLVGVRHASPEDIKAYRKIALKWHDPKNPENKEAEARKKQVAEALEYLSDAK 60
Db 1 MVDYEVLVGVRHASPEDIKAYRKIALKWHDPKNPENKEAEARKKQVAEALEYLSDAK 60
QY 61 KRDIYDKYKGEGLNGGGGSHFDSPEFEFTFRNPDVFRFEGGRDPFSDFEDDPE 120
Db 61 KRDIYDKYKGEGLNGGGGSHFDSPEFEFTFRNPDVFRFEGGRDPFSDFEDDPE 120
QY 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
Db 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
QY 181 GSGMGNFKSIISTYKMNKRIITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
Db 181 GSGMGNFKSIISTYKMNKRIITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
QY 241 ERMRRGQNVLPAPAGLRPPKPPRPSLIRHXPHCLSKKEEGODRPMAPXXMPLASXAG 300
Db 241 ERMRRGQNVLPAPAGLRPPKPPRPSLIRHXPHCLSKKEEGODRPMAPXXMPLASXAG 300
QY 301 XXEGXKRMXAERGVVEEVDQROSLDRT 330
Db 301 XXEGXKRMXAERGVVEEVDQROSLDRT 330
```

```
RESULT 3
US-09-388-993-3
; Sequence 3, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAM-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/668,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-09-388-993-3
```

Query Match 98.8%; Score 1696; DB 3; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.9e-167;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPNKEEAEKKFKQVAEAYEVLSDAK 60
DB 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPNKEEAEKKFKQVAEAYEVLSDAK 60
QY 61 KRDIYDKYKGBGLNGGGGGSHEDSPFEFGTFRNPDVREFFGADPFSEDFEEDPFE 120
DB 61 KRDIYDKYKGBGLNGGGGGSHEDSPFEFGTFRNPDVREFFGADPFSEDFEEDPFE 120
QY 121 DFGNRRGPGSRGSRGSGFSAFSGPSSGFSFDGTGTSGLGHGGLTSFSSSTSF 180
DB 121 DFGNRRGPGSRGSRGSGFSAFSGPSSGFSFDGTGTSGLGHGGLTSFSSSTSF 180
QY 181 GGSAGMNFKSIISTKKNVNGRKITTKRIVENGOEVEVEEDGQLKSLTINGVADDALXE 240
DB 181 GGSAGMNFKSIISTKKNVNGRKITTKRIVENGOEVEVEEDGQLKSLTINGVADDALXE 240
QY 241 ERMRRGQNVLPAPAGLRPPAPASLRLHXPHCLSKKEGQDRPMAPIXXWPLASXAG 300
DB 241 ERMRRGQNVLPAPAGLRPPAPASLRLHXPHCLSKKEGQDRPMAPIXXWPLASXAG 300
QY 301 XXEGXKRMXAEAEKGVVEEVDQROSLDRT 330
DB 301 XXEGXKRMXAEAEKGVVEEVDQROSLDRT 330

RESULT 4

US-08-868-288A-7
Sequence 7, Application US/08868288A
Patent No. 5922567

GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,288A
FILING DATE: June 3, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank

CLONE: 32470
US-08-868-288A-7

Query Match 37.0%; Score 635.5; DB 2; Length 277;
Best Local Similarity 54.4%; Pred. No. 7.2e-58;
Matches 142; Conservative 30; Mismatches 48; Indels 41; Gaps 9;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPNKEEAEKKFKQVAEAYEVLSDAK 60
DB 1 MASYEILIDVRSASADIKAYRKALKWHPDKNPNKEEAEKKFKQVAEAYEVLSDAK 60
QY 61 KRDIYDKYKGBGLNGGGGGSHEDSPFEFGTFRNPDVREFFGADPFSEDFEEDPFE 117
DB 61 KREIYDRGRGLTGTGTPRAEAGSGGPGFTFNSPEVEVEEFGSDPFA-ELFDD 119
QY 118 --PEDEFGNRRGPGSRGSRGSGFSAFSGPSSGFSFDGTGTSGLGHGGLTSF 175
DB 120 LGPSEL--QNRG-----SRHSGPFTTSSSP-----GH--SDF 150
QY 176 SSTSFGGS--GKNFKSISTKKNVNGRKITTKRIVENGOEVEVEEDGQLKSLTINGVAD 234
DB 151 SSSFSFSGAGARFVSSTTFVQGRITTRIMENGOEVEVEEDGQLKSLTINGVAD 210
QY 235 DDALXEERMRGQNVLPAPQA 255
DB 211 DLARGELESRRO----QPS 226

RESULT 5

US-09-235-373-7
Sequence 7, Application US/09235373
Patent No. 6001598

GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,373
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/868,288
FILING DATE: June 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 32470

US-09-235-373-7

Query Match	37.0%;	Score 635.5;	DB 3;	Length 277;
Best Local Similarity	54.4%;	Pred. No. 7.2e-58;		
Matches 142;	Conservative 30;	Mismatches 48;	Indels 41;	Gaps 9

```

QY      1 MVDYVEIVGVORHNSPEEDIKKAYKRLILKMHPPDKNPNKNEAEKKKQVAAEYVUSDAK 60
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1 MASYEILIDVFRSASADDIKKAIYRKALOMHPDKNPDKKEAEKKKEVEAEAEVUSDKH 60
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      61 KRDIYDKYKKEGLNGGGGGGSHFDSF-- EFGFTFRNPDPVFEKEEFGAGDPFSDFEED 117
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      61 KREIYDRGREGELGTGTGTPSRAAGGGGPGFTTFFSPSEVFEFEFGSDPFA-ELFDD 119
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      118 --PEDEFFGNRGGRGSRNRGTSGFFSAFSGFPBFGSGFSFSDTGTFFSGLHGCLTIF 175
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      120 LGPFSSEL--QNRG-----SRHSGPFFTFSSSF-----GH--SDF 150
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      176 SSTSFGGS-GGNGFNKSISTSTKMWNGRRITTKRIYVNGQERVEEEDGOLKSLTINGVAD 234
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      151 SSSFSFSPGAGARVYSTSTTFVQGRITTRRMENGGQERVEEEDGOLKSLTINGVPD 210
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      235 DDALXERMRRGQNVLPACQA 255
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      211 DLARGLELSTRREQ-----QPS 226
      1 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 6

US-09-388-993-7
; Sequence 7, Application US/09388993
; Patent No. 604322

```

? GENERAL INFORMATION:
? APPLICANT: Au-Young, Janice
? APPLICANT: Lal, Preeti
? APPLICANT: Bandman, Olga
? TITLE OF INVENTION: TWO NEW HUMAN DNA-J-LIKE PROTEINS
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:

```

Query Match	37.0%;	Score 635.5;	DB 3;	Length 277;
Best Local Similarity	54.4%;	Pred. No. 7.2e-58;		
Matches 142;	Conservative 30;	Mismatches 48;	Indels 41;	Gaps 9;

```

QY 1 HVDYEVGVORHNSPEDIKKATYKRLAKMPDKNPNKEAEKRFQVAEYVLSDAK 60
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1 MASYEILDVPRNSADDIKKATYRKRLAQHMPDKNPNPDKKEAEKKEFEVAEAYVLSDKH 60
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 61 KRDIYDKGKGLNGGGGGGSHDPSF---EFGTFRNPDVDVREFFGNDPSEDFEED 117
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 61 KREIYDRGREGILGTGTGSPRAAGSGCGPDTFFPSPEVEFRREFGSDPFA-ELTFD 119
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 118 --PEDFPGNCRGPRGSRGTSRGTSGFFSAFSGFPSTFGSGSFDTGTFSTGSLGGGLTF 175
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 120 LGPSEEL-QMKG-----SRHSQFFTFPSSSF-----GH---SDF 150
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 176 SSTSPGGS-GGNFKSISTSTKMYNGKRTTKRIIVENGQERVEYEEDGQLKSLTINGVAD 234
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 151 SSSFSFSPGGAARSVSTSTTFQOGRRTTRRIMENGQERVEYEEDGQLKSVTINGVPD 210
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 235 DDALEEEMRRGQNVLPAAQA 255
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 211 DLARGLLELRREQ-----QPS 226
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

RESULT 7
HE-08-069

US-08-868-288A-6
; Sequence 6, Application US/08868288A
: Patent No. 5932567

```

? GENERAL INFORMATION:
? APPLICANT: Au-Young, Janice
? APPLICANT: Lal, Preeti
? APPLICANT: Bandman, Olga
? TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
?

```

Query Match	37.0%;	Score 635.5;	DB 2;	Length 351;
Best Local Similarity	54.4%;	Pred. No. 1e-57;		
Matches 142;	Conservative	30;	Mismatches 48;	Indels 41;
				Gaps 9

[illegible]

RESULT 8
 US-09-235-373-6
 Sequence 6, Application US/09235373
 Patent No. 6001598
 GENERAL INFORMATION:
 APPLICANT: Au-Young, Janice
 APPLICANT: Lal, Preeti
 APPLICANT: Bandman, Olga
 TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/235,373
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/868,288
 FILING DATE: June 3, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0309 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-845-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 351 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: Genbank
 CLONE: 32469
 US-09-235-373-6

Query Match	37.0%;	Score 635.5;	DB 3;	Length 351;
Best Local Similarity	54.4%;	Pred. No. 1e-57;		
Matches 142;	Conservative 30;	Mismatches 48;	Indels 41;	Gaps 9

QY	1	MVDYEVGVORHNSP	EDIKAYAKRL	LKMPDKNPENKKEAEKFFOVAAEAVLS	DAK	60
Db	1	MASYEIIIDVRSAS	ADDIKAIYRKALQ	HPDKNPDKNEKFAEKKFEVAAEAVLS	DKH	60
QY	61	KRDIDYKXGKGLNG	GGGGGGSHDSPF	-	EFGFTFNNPDVFEREFPGADPFS	DFEED 117
Db	61	KREIYDRRGREG	LGTGTGGPRAAG	SGGPGFTTFFPS	PEVEFEFGSDPFA	-ELFDD 119
QY	118	--PEDPFGNRRGR	SGHSRGTSGFFA	SGCFPSFGSGFS	SFDIGTGFSPGSLGAGL	TSF 175
Db	120	LGPSEL--QNRG	-----SRHSGPFFPSSP	-----	-----GH--	SDF 150
QY	176	SSTFSFGGS-GMG	NFKSISTSTKMYN	KRIITKRIYVNGOE	REVEEDGQLKSLT	INGVAD 234
Db	151	SSSSFSFGAGAR	SVSTSTFTVQGR	ITTRIMENGOE	REVEEDGQLKSVT	INGVPD 210
QY	235	DDALKEEMRRGQ	NVLAQPA	235		
Db	211	DLARGLERREO	-----QPS	226		

RESULT 9
 US-09-388-993-6
 Sequence 6, Application US/09388993
 Patent No. 6043222
 GENERAL INFORMATION:
 APPLICANT: Au-Young, Janice
 APPLICANT: Lal, Preeti
 APPLICANT: Bandman, Olga
 TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/388,993
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/868,288
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0309 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-845-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 351 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: Genbank
 CLONE: 32469
 US-09-388-993-6

Query Match 37.08; Score 635.5; DB 3; Length 351;

Best Local Similarity 54.4%; Pred. No. 1e-57;
Matches 142; Conservative 30; Mismatches 48; Indels 41; Gaps 9;

QY 1 MVDYEVYLGVRHASPEDIKAYKRLAKMHPDKNPENKEAEKRFQVAEAYVLSDAK 60
Db 1 MASTYELDPVRSASADIKKAYRKRLAKMHPDKNPENKEAEKRFQVAEAYVLSDAK 60
QY 61 KRDIYDKYEGKINGGGGSHFDSPP---EFQTFRNPDVFEFGGSDPFSDFEDF 117
Db 61 KREIYDYGREGGLTGTGTSRAEAGSGGFTFTFRSPDEVPFRFGSDPFA-ELFDD 119
QY 118 --FPEDFGNRGPRGSRSGTGSFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGF 175
Db 120 LGPFSEL--QNRG-----SRHSGPFFTFSSSF-----GH---SDF 150
QY 176 SPSFSGS--GMGNFKSISTSTKMNNGRKITTKRIYENGQERVEVEDGOLKSLTINGVAD 234
Db 151 SSSSFSFSPAGAFRSYSTITTVQGRITTRIMENGQERVEVEDGOLKSLTINGVAD 210
QY 235 DDALXERMRGQNVLPQPA 255
Db 211 DLARGLELSRREQ-----QPS 226

RESULT 10
US-08-974-546-5
; Sequence 5, Application US/08974546
; Patent No. 5945287

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Corley, Neil C.

APPLICANT: Shah, Puryi

TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,546

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0428

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 340 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 1816452

US-08-974-546-5

Query Match 21.2%; Score 364.5; DB 2; Length 340;

Best Local Similarity 32.9%; Pred. No. 9.2e-30;
Matches 99; Conservative 46; Mismatches 81; Indels 75; Gaps 13;

QY 3 DYIEVLGVORHASPEDIKAYKRLAKMHPDKNPENKEAEKRFQVAEAYVLSDAK 62
Db 4 DYIOTGLAGASADEIKKAYRKRLAKMHPDKNPENKEAEKRFQVAEAYVLSDAK 61
QY 63 DIYDKYEGKINGGGGSHFDSPPFEFGTFR-NPDVFEFGGSDPFSDFEDF 116
Db 62 EIDRIGEBGLKSGSGSGGANGTS--SYTHGDHAFAEFGGR----- 110
QY 117 DPEDFEFGNRGPRGSRSGTGSFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGF 176
Db 110 NPEDTFEGQNRGEGM-----DIDDPFGSP-----MGMG--FT 142
QY 177 STFGSGSGMGNFKSISTSTKMNNGRKITTKRIYENGQERVEVE--DGOLKSLTIN- 231
Db 143 NVNFGRS-----RSAQEPARKKODPPVT-----HDLRVSLSEIYSGCTKKMKISHKRL 190
QY 231 ----GVADDDALXERMRG-----QNVLPQPAQLRPPKPPRPSILRHXPCHLSKEE 280
Db 191 NPDKSIRNEDKILTEVKGKREGKITTPKGGDOTSNMIPADIVLKDHPNLFKRD 250
QY 281 G 281
Db 251 G 251

RESULT 11
US-08-974-546-1
; Sequence 1, Application US/08974546
; Patent No. 5945287

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Corley, Neil C.

APPLICANT: Shah, Puryi

TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,546

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0428

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 348 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: BRAITUT21

CLONE: 2525691

US-09-235-373-5

```

; CLONE: 3
ms-09-060-200-5

```


PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/868,288
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0309 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 397 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 306714
 US-09-388-993-5

```

? FILING DATE: 26-JUL-1996
? CLASSIFICATION: 800
? ATTORNEY/AGENT INFORMATION:
? NAME: Murashige, Kate H.
? REGISTRATION NUMBER: 29,959
? REFERENCE/DOCKET NUMBER: 27112-20038.000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 887-1500
? TELEFAX: (202) 822-0168
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 419 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? OS-08-686-417-3

```

Query Match	16.0%;	Score 274.5;	DB 2;	Length 419;
Best Local Similarity	45.9%;	Pred. No. 2.4e-20;		
Matches 62;	Conservative 15;	Mismatches 23;	Indels 35;	Gaps 5;

Sat Jul 1 21:34:15 2000

us-09-501-714-3.rai

Page 9

[illegible]

Search completed: June 30, 2000, 09:57:48
Job time: 2668 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:57:22 ; Search time 20.55 Seconds

(without alignments)
380.361 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716
Sequence: 1 MVDYEVGVGRHASPEDK.....EAERGVREHVDQROSLDKT 330

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1696	98.8	330	W94066	Human DnaJ-like pr
2	291	17.0	369	W98475	H. pylori GHPO 885
3	274.5	16.0	419	W00936	Maize DnaJ-related
4	274.5	16.0	419	W83397	Maize DnaJ clone p
5	272	15.9	223	W95710	Homo sapiens fetal
6	259.5	15.1	352	W22358	S. pneumoniae N-te
7	257	15.0	358	W94065	Human DnaJ-like pr
8	240	14.0	415	Y07061	Renal cancer assoc
9	232.5	13.5	438	W59132	Mus musculus Tubb I
10	227.5	13.3	223	R90680	Mouse cysteine str
11	218.5	12.7	504	W36140	Bovine P58 protein
12	217	12.6	288	W98446	H. pylori GHPO 542
13	216	12.6	275	Y11104	H. pylori ORF 05ce
14	216	12.6	278	Y11105	H. pylori ORF hp7e
15	176	10.3	159	W44076	Human secreted pro
16	176	10.3	159	W27640	Secreted protein A
17	127.5	7.4	241	W87996	A human MCG18 prot
18	123.5	7.2	1073	P60569	Sequence of the Ri
19	122.5	7.1	265	W74905	Human secreted pro
20	114.5	6.7	621	Y07029	Breast cancer asso
21	104.5	6.1	320	W55828	Human heterogeneous
22	103.5	6.0	341	W26553	Human heterogeneous
23	103.5	6.0	353	W26552	Human heterogeneous
24	103.5	6.0	353	W54362	Heterogeneous nucl
25	103.5	6.0	353	W50921	Amino acid sequenc
26	103.5	6.0	433	W30256	Zuotin. Membranes
27	100.5	5.9	537	R73991	Bovine oviduct spe
28	100	5.8	668	W55483	H. pylori ORF 14ap
29	100	5.8	677	W55328	H. pylori ORF hp3e
30	98	5.7	707	R79912	Human nucleolin. A
31	98	5.7	707	W84052	Human V3 loop HIV
32	96.5	5.6	643	R60020	Fibronectin. New h
33	93.5	5.4	955	R42088	Human p50 or KBP1
34	93.5	5.4	955	R42234	Human p50 protein

35	93.5	5.4	955	1	R42235	Human p50 protein
36	93.5	5.4	955	1	R42236	Human p50 protein
37	93.5	5.4	955	1	R42237	Human p50 protein
38	93.5	5.4	955	1	R42238	Human p50 protein
39	93.5	5.4	955	1	R42239	Human p50 protein
40	93.5	5.4	955	1	R42240	Human p50 protein
41	93.5	5.4	955	1	R42241	Human p50 protein
42	93.5	5.4	955	1	R42242	Human p50 protein
43	93.5	5.4	955	1	R42243	Human p50 protein
44	93.5	5.4	955	1	R42244	Human p50 protein
45	93.5	5.4	955	1	R42245	Human p50 protein

ALIGNMENTS

RESULT	ID	Score	Location/Qualifiers
1	W94066	1696	
W94066	standard; protein; 330 AA.		
AC	09-APR-1999 (first entry)		
DT	Human DnaJ-like protein, HSPJ2.		
DE	DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukemia;		
KW	immune disorder; inflammation; tissue damage; diabetes; wound healing;		
KW	chromosome mapping.		
OS	Homo sapiens.		
FT	Key		
FT	Misc_difference 239		
FT	/label= unknown		
FT	/note= "encoded by GST"		
FT	Misc_difference 272		
FT	/label= unknown		
FT	/note= "encoded by GNG"		
FT	Misc_difference 290		
FT	/label= unknown		
FT	/note= "encoded by GNG"		
FT	Misc_difference 291		
FT	/label= unknown		
FT	/note= "encoded by NCC"		
FT	Misc_difference 293		
FT	/label= unknown		
FT	/note= "encoded by NNC"		
FT	Misc_difference 298		
FT	/label= unknown		
FT	/note= "encoded by NCA"		
FT	Misc_difference 301		
FT	/label= unknown		
FT	/note= "encoded by TTN"		
FT	Misc_difference 302		
FT	/label= unknown		
FT	/note= "encoded by NAA"		
FT	Misc_difference 305		
FT	/label= unknown		
FT	/note= "encoded by NGC"		
FT	Misc_difference 309		
FT	/label= unknown		
FT	/note= "encoded by NAA"		
FT	W0985509-A2.		
PD	10-DEC-1998.		
PD	02-JUN-1998; U11182.		
PR	03-JUN-1997; US-868288.		
PA	(INCY-) INCYTE PHARM INC.		
PI	Au-Young J, Bandman O, Lal P;		
DR	WPI; 99-070259/06.		
DR	N-PSDB: X06100.		
PT	New nucleic acid encoding human DnaJ-like proteins - for diagnosis,		
PT	treatment and prevention of cancer, immune disorders and		
PT	Inflammation		
PS	Claim 22; Fig 3A-D; 73pp; English.		
CC	This represents a human DnaJ-like protein, HSPJ2. The invention provides		
CC	two human DnaJ-like proteins which are heat shock proteins j1 and j2		
CC	(HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host		
CC	cells containing a vector comprising the nucleic acids are used for the		

CC production of recombinant proteins. Recombinant HSP71 and HSP72 are used
CC to raise Ab, therapeutically and to screen for specific binding agents.
CC Antagonists are used to treat or prevent a wide variety of solid cancers,
CC leukaemia and lymphoma; immune disorders (typical of many disclosed are
CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
CC and parasitic infections) and inflammation. Agonists may be used to treat
CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound
CC healing, and may be expressed from a vector containing the nucleic acids.
CC Fragments of the nucleic acids are used as primers and probes for
CC detecting and quantifying the HSP7-encoding nucleic acid in usual
CC hybridisation and/or amplification assays, therapeutically as antisense,
CC triplex-forming or ribozyme molecules, and for chromosome mapping.
SQ Sequence 330 AA;

Query Match 98.8%; Score 1696; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.2e-165;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDVYEVLYGVRHASPEDIKKAYRKALKMHPDKPENKEAEKRFQVAEAYVLSDAK 60
DB 1 MDVYEVLYGVRHASPEDIKKAYRKALKMHPDKPENKEAEKRFQVAEAYVLSDAK 60
QY 61 KRDIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDVREFFGGRDPSFDFEDPFE 120
DB 61 KRDIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDVREFFGGRDPSFDFEDPFE 120
QY 121 DFGNRRGPRGSRGSGRGSFSAFSGPFGSGFSDTGTFTSGTSGHGLTFSSTSF 180
DB 121 DFGNRRGPRGSRGSGRGSFSAFSGPFGSGFSDTGTFTSGTSGHGLTFSSTSF 180
QY 181 GSGGMGFKISTSTKVNKRKITTKRIYVNGGGRVEEDGOLKLTINGVADDLAXE 240
DB 181 GSGGMGFKISTSTKVNKRKITTKRIYVNGGGRVEEDGOLKLTINGVADDLAXE 240
QY 241 ERMRRGQNVLPAPAGLPPKPPRPASLLRHXPHCLSKKEGEDRPWAPYXWXPPLASXAG 300
DB 241 ERMRRGQNVLPAPAGLPPKPPRPASLLRHXPHCLSKKEGEDRPWAPYXWXPPLASXAG 300
QY 301 XHEGXRMAEAERGVVEEVDROSIDRT 330
DB 301 XHEGXRMAEAERGVVEEVDROSIDRT 330

RESULT 2
W98475
ID W98475 standard; Protein: 369 AA.

AC W98475; 31-MAR-1999 (first entry)
DE H. pylori GHPD 885 protein.
KW K61 protein; Helicobacter infection; gastroduodenal disease; gastritis;
OS peptic ulcer disease.
KM Helicobacter pylori.
PN W09843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INNR) MERIEUX OCAVAX PASTEUR MERIEUX SERONS.
PI A1-garw1 A, Kleuthous H, Miller C, Oomen RP, Tomb J;
DR WPI: 98-542293/46.
DR N-PSDB: X14194.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 8; Page 853-855; 2054pp; English.
CC This sequence represents a Helicobacter pylori GHPD protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with

CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 369 AA;

Query Match 17.0%; Score 291; DB 1; Length 369;
Best Local Similarity 32.0%; Pred. No. 1e-21;
Matches 93; Conservative 32; Mismatches 80; Indels 86; Gaps 12;

QY 2 VDYEVLYGVRHASPEDIKKAYRKALKMHPDKPENKEAEKRFQVAEAYVLSDAK 61
DB 3 LSYELLEVAKHSHNOETIKSYRKALKYIPDNAGDK-EAEKRFQVAEAYVLSDAK 61
QY 62 RDYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDVREFFGGRDPSFDFEDPFE 119
DB 62 RALYDRGKKGKLNQAGASQDGS-----DFEEDLGSF 93
QY 119 FEFDFGRRGPRGSRGSGT-----SFGSAPSGFP-----SFGSGRSSFD-TGF 161
DB 94 FEDAFG--FGARSKKQKSSADPVLQTLSEKFAVFGCKTKIKYQYVSCSCDGTGA 151
QY 162 -----TSGSLGSG-----GLTSFSTSGSGGMGFKISTSTKVNKRKITTKRIY 209
DB 152 KDKALEFCQCNQGVFMNQGFMSFAQTC--CAGCGKRIYVTPQACKGKIYILK--- 207
QY 210 ENGQERVEVEDGOLKLTINGVADDL-----XEMRRGQNVLPAP 253
DB 207 -----DEIDALIPGIDQDNRMYLKNKGNEYKRGKGLYLEAQ 246

RESULT 3

W00936
ID W00936 standard; Protein: 419 AA.

AC W00936; 11-NOV-1997 (first entry)
DE Maize DnaJ-related protein.
KW DnaJ-related protein; chaperone; protein folding; DNA replication;
KW translation; peptide translocation; zmdJ1; promoter; maize;
KW transgenic plant; insecticide; antifungal; fungicide;
KW crop protection.
OS Zea mays L. cv. B73.
PN W09705260-A2.
PD 13-FEB-1997.
PF 12-JUL-1996; U11676.
PR 26-JUL-1995; US-001522.
PA (PION-) PIONEER HI-BRED INT INC.
PI Barbour E, Baszczynski C, Horowitz J, Kosichan JL;
DR WPI: 97-145697/13.
DR N-PSDB: T84330.
PT Tobacco DnaJ-related gene transcription/translation regulatory
PT sequence, zmdJ1 - is intermediate between constitutive and tissue
PT specific promoters, partic. for control of antifungal and
PT insecticidal genes
PS Example 1; Fig 2A-B; 26pp; English.
CC This polypeptide sequence is encoded by the maize DnaJ-related
CC gene (see T84330). DnaJ-related proteins assist in chaperone-
CC mediated protein folding and provide cell viability at high
CC temperatures. They are also involved in DNA replication,
CC translation and peptide translocation across intracellular
CC membranes. Due to this wide range of functions, DnaJ has a wide
CC range of effectiveness and the gene's promoter sequence, zmdJ1
CC (see T84330), is effective in a wide range of tissues. The
CC claimed zmdJ1 promoter can be used in claimed methods for control
CC of antifungal or insecticidal genes in transgenic plants.
SQ Sequence 419 AA;

Query Match 16.0%; Score 274.5; DB 1; Length 419;
Best Local Similarity 45.9%; Pred. No. 6.1e-20;
Matches 62; Conservative 15; Mismatches 23; Indels 35; Gaps 5;

PF 17-MAY-1996; CA0322.
 PR 07-JUN-1995; US-472534.
 PR 04-AUG-1995; US-001805.
 PA (IAFBI) IAF BIOVAC INC.
 PI Brodeur B, Hamel J, Martin D, Rioux C;
 DR WPI: 97-052328/05.
 DR N-PSDB: T73388.
 PT Streptococcal heat shock proteins and corresponding DNA sequences -
 PT used in the production of a vaccine to treating and preventing
 PT strain-specific Streptococcal infection
 PS Example 3: Page 93-94: 156pp: English.
 CC This amino acid sequence corresponds to the N-terminal portion of the
 CC Streptococcus pneumoniae DnaJ protein. The protein shows a high degree
 CC of identity (72% and 51%) to the DnaJ proteins from *Lactococcus lactis*
 CC and *E. coli* respectively. The sequence encoding this truncated protein
 CC was isolated on the same nucleotide sequence that encodes the
 CC S. pneumoniae heat shock protein 72 (HSP72; W2357). The nucleotide
 CC fragment was isolated from a HindIII-partially digested genomic DNA
 CC library using a fragment of the chimeric gene (T73392) corresponding to
 CC the sequence encoding the C-terminal 169 amino acids of HSP72. The HSP72
 CC protein and its fragment, or antibodies specific to HSP72, are used in
 CC pharmaceutical compositions, pref. a vaccine, for treating or preventing
 CC infection by S. pneumoniae or related bacteria in humans, e.g.
 CC S. pyogenes or S. agalactiae.
 SQ Sequence 352 AA;

Query Match 15.1%; Score 259.5; DB 1; Length 352;
 Best Local Similarity 32.7%; Pred. No. 1.6e-18;
 Matches 81; Conservative 30; Mismatches 82; Indels 55; Gaps 10;

QY 3 DYEVYGVQNHASPEDIKKAYRKALKMHPDKNPENKEAEKFKOVAEAYEVLSDAKKR 62
 DB 5 EYDRLGVSNMAADEIKKAYRKALKMHPDKNPENKEAEKFKOVAEAYEVLSDAKKR 62
 QY 63 DIYDKYKESGLNGGGGSHFSDPFEGFTFRNPDVFRFEFGS-----RPP----- 110
 DB 63 AADYQAGAGANGFGGAGGFGG-FNGAGGFGFEDIFFSFGGSSRNPNAPRGDDI 121
 QY 110 -----FSPDFEEDPFDFEGNRR-----GPRGSRSGTGSFSAFSGFSGSGSS 156
 DB 122 QYRVNLTFF-----EAFETKEVYKVRHAGCRTCNGSAGKPGTSPVTCGRCHGAGVYN 175
 QY 157 FDDGFTSFGSL-----GHGGLTSFSSTSGSGMGNGKESISTSKMANGKRIITTK 206
 DB 176 VDTQ-TPLGMRQVTCVCHGKKEIKYPCCTTGCTGHEK-QAHSVHVKIPAG----- 228
 QY 207 RIVENGQOE 214
 DB 228 --VETGQO 233

RESULT 7
 ID W94065 standard; Protein: 358 AA.
 AC W94065;
 DT 09-APR-1999 (first entry)
 DE Human DnaJ-like protein, HSPJ1.
 KW DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;
 KW immune disorder; inflammation; tissue damage; diabetes; wound healing;
 KW chromosome mapping.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT MISC-difference 298
 FT //label= unknown
 FT //note= "encoded by AAN"
 FT MISC-difference 299
 FT //label= unknown
 FT //note= "encoded by TAN"
 PV W09855509-A2.
 PD 10-DEC-1998.
 PF 02-JUN-1998; U11182.
 PR 03-JUN-1997; US-868288.

PA (INCY-) INCYTE PHARM INC.
 PI Al-Young J, Bandman O, Tai P;
 DR WPI: 99-070259/06.
 DR N-PSDB: X06099.
 PT New nucleic acid encoding human DnaJ-like proteins - for diagnosis,
 PT treatment and prevention of cancer, immune disorders and
 PT inflammation
 PS Claim 1: Fig 1A-D, 73pp, English.
 CC This represents a human DnaJ-like protein, HSPJ1. The invention provides
 CC two human DnaJ-like proteins which are heat shock proteins J1 and J2
 CC (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host
 CC cells containing a vector comprising the nucleic acids are used for the
 CC production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used
 CC to raise Ab, therapeutically and to screen for specific binding agents.
 CC Antagonists are used to treat or prevent a wide variety of solid cancers,
 CC leukaemia and lymphoma; immune disorders (typical of many disclosed are
 CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
 CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
 CC and parasitic infections) and inflammation. Agonists may be used to treat
 CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
 CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound
 CC healing, and may be expressed from a vector containing the nucleic acids.
 CC Fragments of the nucleic acids are used as primers and probes for
 CC detecting and quantifying the HSPJ-encoding nucleic acid in usual
 CC hybridisation and/or amplification assays, therapeutically as antisense,
 CC triplex-forming or ribozyme molecules, and for chromosome mapping.
 SQ Sequence 358 AA;

Query Match 15.0%; Score 257; DB 1; Length 358;
 Best Local Similarity 46.2%; Pred. No. 3e-18;
 Matches 60; Conservative 18; Mismatches 30; Indels 22; Gaps 5;

QY 3 DYEVYGVQNHASPEDIKKAYRKALKMHPDKNPENKEAEKFKOVAEAYEVLSDAKKR 62
 DB 25 DYKTLGVRSASIKDIAKAYRKALKMHPDKNPDPD-PQAEKFFODLAAEYVLSDSSEKR 83
 QY 63 DIYDKYKESGLNGGGGSHFSDPFEGFTFRNPDVFRFEFGSDPSFDFEEDPFDFE 122
 DB 84 KQDYDYGEGGLK-DGHQSH-----GDITSHFFG---DFGFMGGTIPRQO- 125
 QY 123 FGNRRGPRGS 132
 DB 125 --DRNIPRGS 132

RESULT 8
 ID Y07061 standard; Protein: 415 AA.
 AC Y07061;
 DT 02-JUL-1999 (first entry)
 DE Renal cancer associated antigen precursor sequence.
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 OS Homo sapiens.
 PN W09904265-A2.
 PD 28-JAN-1999.
 PF 15-JUL-1998; U14679.
 PR 22-JUN-1998; U14679.
 PR 17-JUL-1997; US-896164.
 PR 10-OCT-1997; US-061599.
 PR 10-OCT-1997; US-061765.
 PR 10-OCT-1997; US-948705.
 PR 11-OCT-1997; GB-021697.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Chen Y, Cout I, Ohare A, Odata Y, Old LJ,
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
 PI Tureci O;
 DR WPI: 99-132448/11.
 PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers

PS Claims 37, 41; Page 270-271; 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 275 AA;

Query Match 12.6%; Score 216; DB 1; Length 275;
 Best Local Similarity 29.2%; Pred. No. 3.2e-14;
 Matches 66; Conservative 30; Mismatches 70; Indels 60; Gaps 9;

QY 5 YEVLGVRHASPEDIKKAYRKALKMHPDNPKNEKEAEKFKQVAAVEYLSDAKKRDI 64
 DB 6 YQTLNVSNNASQDEIKKSYRLANQYHPDLN--KTKAEAEKFEINAAVEYLSDEEKRRQ 63
 QY 65 YDKYKGEGLNGGGGGSHFDSPEFGFTFRNPDDVREFRFGGRDPFSFDFEDDFEDFFG 124
 DB 64 YDQFGD-----NMFGGN-----FSD-FA 81
 QY 125 NRRGPRGSRGRTGTSFSAFGSPFGSGFSFDGTGTSGLGHC---GLTFSFSSTSF 180
 DB 82 RSRGP-----SEDDLDLISLTFKGGFSQRFSGFNSGFNFAPENDVTAIILNVS 137
 QY 181 GSGMGNGFKSISTSKMNGRKITTKRI--VENGOEVEVEEDGQL 224
 DB 138 LDTLLGNKKQV-----VNNETPSLKIPIGVEBG-EKIRVRNKGKM 177

RESULT 14
 Y11105
 ID Y11105 standard; Protein; 278 AA.
 AC Y11105;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF hp7e10192.23712780.f2.5 cytoplasmic protein.
 KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KM secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN W09824475-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997; U22104.
 PR 14-JUL-1997; US-891928.
 PR 05-DEC-1996; US-759625.
 PR 25-MAR-1997; US-823745.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 DR WPI: 98-333051/29.
 DR N-PSDB: X30634.
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PS Claims 37, 41; Page 271-272; 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 278 AA;

Query Match 12.6%; Score 216; DB 1; Length 278;
 Best Local Similarity 29.2%; Pred. No. 3.3e-14;
 Matches 66; Conservative 30; Mismatches 70; Indels 60; Gaps 9;

QY 5 YEVLGVRHASPEDIKKAYRKALKMHPDNPKNEKEAEKFKQVAAVEYLSDAKKRDI 64
 DB 6 YQTLNVSNNASQDEIKKSYRLANQYHPDLN--KTKAEAEKFEINAAVEYLSDEEKRRQ 63
 QY 65 YDKYKGEGLNGGGGGSHFDSPEFGFTFRNPDDVREFRFGGRDPFSFDFEDDFEDFFG 124
 DB 64 YDQFGD-----NMFGGN-----FSD-FA 81
 QY 125 NRRGPRGSRGRTGTSFSAFGSPFGSGFSFDGTGTSGLGHC---GLTFSFSSTSF 180
 DB 82 RSRGP-----SEDDLDLISLTFKGGFSQRFSGFNSGFNFAPENDVTAIILNVS 137
 QY 181 GSGMGNGFKSISTSKMNGRKITTKRI--VENGOEVEVEEDGQL 224
 DB 138 LDTLLGNKKQV-----VNNETPSLKIPIGVEBG-EKIRVRNKGKM 177

RESULT 15
 W44076
 ID W44076 standard; Protein; 159 AA.
 AC W44076;
 DT 12-MAY-1998 (first entry)
 DE Human secreted protein Am610.
 KW Human; secreted protein; ATCC 98026; cytokine; immunomodulation;
 OS cell proliferation; differentiation; regulation.
 PH Homo sapiens.
 FT location/Qualifiers
 FT Misc.-difference 82 /label= Unspecified
 FT /note= "encoded by AGN"
 FT Misc.-difference 87 /label= Unspecified
 FT /note= "encoded by TGA a stop codon"
 FT Misc.-difference 96 /label= Unspecified
 FT /note= "encoded by TAG a stop codon"
 FT Misc.-difference 102 /label= Unspecified
 FT /note= "encoded by RGR"
 FT Misc.-difference 103 /label= Unspecified
 FT /note= "encoded by RGR"
 FT Misc.-difference 106 /note= "encoded by RGT"
 FT /label= "Unspecified
 FT /note= "encoded by RAG"
 FT Misc.-difference 109 /label= "Unspecified
 FT /note= "encoded by TGA a stop codon"
 FT Misc.-difference 110 /label= "Unspecified
 FT /note= "encoded by GYR"
 FT Misc.-difference 111 /label= "Unspecified
 FT /note= "encoded by NKC"
 FT Misc.-difference 112 /label= "Unspecified
 FT /note= "encoded by NKT"
 FT Misc.-difference 113 /label= "Unspecified
 FT /note= "encoded by TMN"
 FT Misc.-difference 114 /label= "Unspecified
 FT /note= "encoded by YTT"
 FT Misc.-difference 115 /label= "Unspecified
 FT /note= "encoded by YAA"
 FT Misc.-difference 116 /label= "Unspecified
 FT /note= "encoded by YTT"

FT Misc_difference 117 /label= Unspecified
 FT /note= "encoded by TSA"
 FT Misc_difference 118 /label= Unspecified
 FT /note= "encoded by TGA a stop codon"
 FT Misc_difference 121 /label= Unspecified
 FT /note= "encoded by TAA a stop codon"
 FT Misc_difference 128 /label= Unspecified
 FT /note= "encoded by TYN"
 FT Misc_difference 129 /label= Unspecified
 FT /note= "encoded by ARR"
 FT Misc_difference 130 /label= Unspecified
 FT /note= "encoded by CVA"
 FT Misc_difference 134 /label= Unspecified
 FT /note= "encoded by AKC"
 FT Misc_difference 135 /label= Unspecified
 FT /note= "encoded by YAA"
 FT Misc_difference 136 /label= Unspecified
 FT /note= "encoded by NAA"
 FT Misc_difference 137 /label= Unspecified
 FT /note= "encoded by YKR"
 FT Misc_difference 139 /label= Unspecified
 FT /note= "encoded by GRP"
 FT Misc_difference 140 /label= Unspecified
 FT /note= "encoded by RWY"
 FT Misc_difference 141 /label= Unspecified
 FT /note= "encoded by CAW"
 FT Misc_difference 142 /label= Unspecified
 FT /note= "encoded by WYC"
 FT Misc_difference 143 /label= Unspecified
 FT /note= "encoded by CNN"
 FT Misc_difference 146 /label= Unspecified
 FT /note= "encoded by NNN"
 FT Misc_difference 147 /label= Unspecified
 FT /note= "encoded by WKG"
 FT Misc_difference 149 /label= Unspecified
 FT /note= "encoded by SYC"
 FT Misc_difference 151 /label= Unspecified
 FT /note= "encoded by GNG"
 FT Misc_difference 155 /label= Unspecified
 FT /note= "encoded by TGN"
 PN W09739123-A2.
 PD 23-OCT-1997.
 PF 14-APR-1997; U06139.
 PR 18-APR-1996; US-634325.
 PA (GENY) GENETICS INST INC.
 PI Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racle LA,
 PI Spaulding V;
 DR WPI: 97-526460/48.
 DR N-PSDB; V02136.
 PR New secreted proteins encoded clones present in ATCC 98026 -
 PR possibly having cytokine, cell proliferation/differentiation
 PR regulating, immunomodulating and many other activities
 PS Claim 12; Page 78-79; 139pp. English.

CC The present sequence represents a novel human secreted protein deposited
 CC under accession number ATCC 98026. The secreted protein can be used to
 CC determine biological activity, to raise antibodies, as tissue markers,
 CC to isolate cognate ligands or receptors, to identify agents that
 CC modulate their interactions and as nutritional supplements. It may also
 CC have a very wide range of biological activities although no evidence
 CC for any is provided in the specification. Typical of these are cytokine,
 CC cell proliferation/differentiation modulating activity or induction of
 CC other cytokines; immunostimulating/immunosuppressant activities (e.g.
 CC for treating human immunodeficiency virus infection, cancer, autoimmune
 CC diseases and allergy); regulation of haematopoiesis (e.g. for treating
 CC anaemia or as adjunct to chemotherapy); stimulation of growth of bone,
 CC cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds,
 CC periodontal disease, neurological diseases stroke, fibrosis); inhibition
 CC or stimulation of follicle stimulating hormone (for control of
 CC fertility); chemotactic and chemokinetic activities (e.g. for treating
 CC infections, tumours); haemostatic or thrombolytic activity (e.g. for
 CC treating haemophilia, cardiac infarction etc.); anti-inflammatory
 CC activity (e.g. for treating septic shock, Crohn's disease); as
 CC antimicrobials; for treating psoriasis or other hyperproliferative
 CC disease; for regulation of metabolism, behaviour, and many others. Also
 CC contemplated is the use of the corresponding nucleic acid in gene
 CC therapy procedures.
 CC
 CC Sequence 159 AA:

Query Match 10.3%; Score 176; DB 1; Length 159;
 Best Local Similarity 56.2%; Pred. No. 1.8e-10;
 Matches 56; Conservative 10; Mismatches 14; Indels 4; Gaps 2;
 QY 4 YVEVLGVORHASPEDIKKAYRLALMKWHPDKNPENKEERFKQYAEAYEVLSDAK--K 61
 Db 27 YYDILGVPRKSASERQIKKAFHKLAMKYHPDKN--KSPDAENAFRELAAYETLSDANXK 84
 QY 62 RDLY 65
 Db 85 RYXY 88

Search completed: June 30, 2000, 09:57:24
 Job time: 2725 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 13:32:40 ; Search time 49.41 Seconds
(without alignments)
3498.889 Million cell updates/sec

Title: US-09-501-714-4

Perfect score: 1330

Sequence: 1 CGNAGAGAGNMAAGGAAG.....TGACGGCAGCGGTGGCGGGG 1330

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 6492525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/1na/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5D_COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/1na/6_COMB.seq:*
6: /cgn2_6/ptodata/2/1na/PCrUS_COMB.seq:*
7: /cgn2_6/ptodata/2/1na/backfilest.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1307.2	98.3	1330	4	Sequence 4, Appli
2	1307.2	98.3	1330	5	Sequence 4, Appli
3	1307.2	98.3	1330	5	Sequence 4, Appli
4	91.4	6.9	1376	4	Sequence 2, Appli
5	91.4	6.9	1376	5	Sequence 2, Appli
6	91.4	6.9	1376	5	Sequence 2, Appli
7	75.6	5.7	672	2	Sequence 2, Appli
8	73.4	5.5	2349	4	Sequence 2, Appli
9	68.2	5.1	4320	4	Sequence 2, Appli
10	66.4	5.0	1756	4	Sequence 4, Appli
11	60	4.5	1515	2	Sequence 3, Appli
12	60	4.5	1687	1	Sequence 14, Appli
13	55.8	4.2	1700	1	Sequence 14, Appli
14	55.2	4.2	1700	1	Sequence 14, Appli
15	54	4.1	3748	3	Sequence 2, Appli
16	34.4	2.6	13987	3	Sequence 13, Appli
17	34.4	2.6	44377	3	Sequence 7, Appli
18	34.4	2.6	44377	3	Sequence 7, Appli
19	34.2	2.6	1704	2	Sequence 5, Appli
20	34.2	2.6	1704	2	Sequence 5, Appli
21	34.2	2.6	1908	2	Sequence 8, Appli
22	34.2	2.6	1908	3	Sequence 8, Appli
23	34	2.6	1787	3	Sequence 8, Appli
24	34	2.6	3211	4	Sequence 8, Appli
25	34	2.6	3901	4	Sequence 8, Appli
26	33.8	2.5	1017	6	Sequence 21, Appli
27	33.4	2.5	1785	5	Sequence 16, Appli

28	33	2.5	1662	2	US-08-671-947-1	Sequence 1, Appli
29	32.6	2.5	1644	1	US-07-903-047-7	Sequence 7, Appli
30	32.6	2.5	3901	1	US-08-188-582-31	Sequence 31, Appli
31	32.6	2.5	3901	1	US-08-646-715-31	Sequence 31, Appli
32	32.6	2.5	44377	3	US-08-804-227C-7	Sequence 7, Appli
33	32.6	2.5	44377	4	US-08-804-198-1	Sequence 1, Appli
34	32.4	2.4	35081	3	US-08-752-760A-1	Sequence 1, Appli
35	31.8	2.4	1055	2	US-08-618-464-1	Sequence 1, Appli
36	31.8	2.4	1055	6	PCT-US95-04896-1	Sequence 1, Appli
37	31.8	2.4	1820	1	US-08-173-508-7	Sequence 7, Appli
38	31.8	2.4	1821	3	US-08-265-310-7	Sequence 7, Appli
39	31.8	2.4	2312	1	US-07-736-178C-1	Sequence 1, Appli
40	31.8	2.4	30001	1	US-08-125-468-1	Sequence 1, Appli
41	31.8	2.4	30001	3	US-08-474-933-1	Sequence 1, Appli
42	31.4	2.4	4852	1	US-07-853-913-3	Sequence 3, Appli
43	31.2	2.3	370	3	US-08-332-766A-8	Sequence 8, Appli
44	31.2	2.3	879	1	US-08-243-545-1	Sequence 1, Appli
45	31.2	2.3	879	3	US-08-993-962-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-868-288A-4
Sequence 4, Application US/08868288A
Patent No. 5922567
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,288A
FILING DATE: June 3, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1330 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HMT2RAT01
CLONE: 260873
US-08-868-288A-4

Query Match 98.3%; Score 1307.2; DB 4; Length 1330;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

|||||
Db 1 CGNAGGAGNNAAGAAAGNCGCCGAGAGCCGCCGNAACCAACGCGNACANTCCT 60
QY 61 GGNCTGTTGAGAGATTGCGGCGCTCACCTGCCCTCCCTGCTCCGCCACGCGCGCT 120
Db 61 GGNCTGTTGAGAGATTGCGGCGCTCACCTGCCCTCCCTGCTCCGCCACGCGCGCT 120
QY 121 TCTTTCTCGGAGCCATTCACAACTCTGTAACATGCTGTAATGCTATGAACTCTA 180
Db 121 TCTTTCTCGGAGCCATTCACAACTCTGTAACATGCTGTAATGCTATGAACTCTA 180
QY 181 GGGGTGAGAGATGCTGCTACCCGAGGATTTTAAAGGATTCGGAACCTGGCACTG 240
Db 181 GGGGTGAGAGATGCTGCTACCCGAGGATTTTAAAGGATTCGGAACCTGGCACTG 240
QY 241 AAGTGGCATCCAGTAAATATCTGAGATTAAGAGAGCAGAGAAAAATTCAGCAA 300
Db 241 AAGTGGCATCCAGTAAATATCTGAGATTAAGAGAGCAGAGAAAAATTCAGCAA 300
QY 301 GTAGCGAGCATTTGAAAGTCTGCGATGCTTAAGAAAGGAGCATCTATGACAAATAT 360
Db 301 GTAGCGAGCATTTGAAAGTCTGCGATGCTTAAGAAAGGAGCATCTATGACAAATAT 360
QY 361 GGCAGAGAGATTTAATGTTGNGNGNGNGTGAAGTCAATTTGACAGTCCATTTGAA 420
Db 361 GGCAGAGAGATTTAATGTTGNGNGNGNGTGAAGTCAATTTGACAGTCCATTTGAA 420
QY 421 TTTGGCTTCACTTCCGTAACCCAGATGATGCTTTCAGGGAAATTTTGTGTAAGGAC 480
Db 421 TTTGGCTTCACTTCCGTAACCCAGATGATGCTTTCAGGGAAATTTTGTGTAAGGAC 480
QY 481 CCATTTTCATTGACTTCTTTGAAGACCCCTTTTGAGACCTTCTTTGGGAATCGAAGGCT 540
Db 481 CCATTTTCATTGACTTCTTTGAAGACCCCTTTTGAGACCTTCTTTGGGAATCGAAGGCT 540
QY 541 CCCCAGAGAGAGAGAGAGAGGAGGAGGCTTTTCTGCTGCTGAGTGGATTTCCG 600
Db 541 CCCCAGAGAGAGAGAGAGAGGAGGAGGCTTTTCTGCTGCTGAGTGGATTTCCG 600
QY 601 TCTTTTGAAGTGAATTTCTTTTGAATACAGATTTACTTCAATTTGGGTCACTAGT 660
Db 601 TCTTTTGAAGTGAATTTCTTTTGAATACAGATTTACTTCAATTTGGGTCACTAGT 660
QY 661 CAGGCGGCGCTCACTTATCTCTCCAGCTCATTTGCTGTAAGTGGCATGGCACTTC 720
Db 661 CAGGCGGCGCTCACTTATCTCTCCAGCTCATTTGCTGTAAGTGGCATGGCACTTC 720
QY 721 AAATGATATCACTTCACTTAAATGTTAATGCGAGAAAAATCAGTACAAAGAGATT 780
Db 721 AAATGATATCACTTCACTTAAATGTTAATGCGAGAAAAATCAGTACAAAGAGATT 780
QY 781 GTGAGAGACGGTCAAGAAAGTAGAAGTTGAAGAAAGATGGCCAGTTAAAGTCTTAA 840
Db 781 GTGAGAGACGGTCAAGAAAGTAGAAGTTGAAGAAAGATGGCCAGTTAAAGTCTTAA 840
QY 841 ATAAATGCTGTGCGCAGCAGATGCCCTGCTGAGAGCGCATGCGAGAGGCCAAGAC 900
Db 841 ATAAATGCTGTGCGCAGCAGATGCCCTGCTGAGAGCGCATGCGAGAGGCCAAGAC 900
QY 901 GTCTGCGACGCCAGCTGCGCGCTGCGAGACCGCGGAAGCCGCCGCTGCTGCTG 960
Db 901 GTCTGCGACGCCAGCTGCGCGCTGCGAGACCGCGGAAGCCGCCGCTGCTGCTG 960
QY 961 CTGAGACACGNGCCTCATTTGCTCTTAAGAGAGAGGCGCAGCAGACCTTGGGCA 1020
Db 961 CTGAGACACGNGCCTCATTTGCTCTTAAGAGAGAGGCGCAGCAGACCTTGGGCA 1020
QY 1021 CCGGNGCCTGAGNCCCTGCTGCTCCNCAAGCAGGNTTNNAGAAGGTNGCAGAGATG 1080
Db 1021 CCGGNGCCTGAGNCCCTGCTGCTCCNCAAGCAGGNTTNNAGAAGGTNGCAGAGATG 1080
QY 1081 NAAGCAGAGAGAGAGAGGTGAGAGAGAGAGAGTCCGACCAAGGCAATCACTAGAC 1140

```

```

Db 1081 NAAGCAGAGAGAGAGAGGTGAGAGAGAGAGAGTCCGACCAAGGCAATCACTAGAC 1140
QY 1141 CGGACTTGAGCAGCGGCTGACACCCAGACGCTGCGGCTCCACGCTGCTGGCATGCG 1200
Db 1141 CGGACTTGAGCAGCGGCTGACACCCAGACGCTGCGGCTCCACGCTGCTGGCATGCG 1200
QY 1201 TCGTGACACGCGCTAGGTAGCAGCGTGCAGTGTCTGTGAGGCGCACATCGCTCG 1260
Db 1201 TCGTGACACGCGCTAGGTAGCAGCGTGCAGTGTCTGTGAGGCGCACATCGCTCG 1260
QY 1261 GCAGATTATGCAATCAGATCAGTCAAGAGCAGAGGTCAGAGAGAGCGGCTGACGCG 1320
Db 1261 GCAGATTATGCAATCAGATCAGTCAAGAGCAGAGGTCAGAGAGAGCGGCTGACGCG 1320
QY 1321 GGTGCGGCGG 1330
Db 1321 GGTGCGGCGG 1330

RESULT 3
US-09-388-993-4
; Sequence 4, Application us/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMT2RAT01
; CLONE: 260873
; US-09-388-993-4

Query Match 98.3%; Score 1307.2; DB 5; Length 1330;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGNAGGAGNNAAGAAAGNCGCCGAGAGCCGCCGNAACCAACGCGNACANTCCT 60
Db 1 CGNAGGAGNNAAGAAAGNCGCCGAGAGCCGCCGNAACCAACGCGNACANTCCT 60

```

Qy	61	GGGCGCTNTGAGGAGATGCGGGGCGGTAACCGGCTCCCGCTCCGCTTCCCGCACCGGGCGCT	120
Db	61	GGNGCTNTGAGGAGATGCGGGGCGGTAACCGGCTCCCGCTCCCGCTTCCCGCACCGGGCGCT	120
Qy	121	TCCTTCTCTGGAGCCCATTCGTAACAATCTCGTAACAAACATGCTGTAATTAATCTATGAAGTTCTA	180
Db	121	TCCTTCTCTGGAGCCCATTCGTAACAATCTCGTAACAAACATGCTGTAATTAATCTATGAAGTTCTA	180
Qy	181	GGGCTGCAAGACATGCTTCACCCGAGGATTTTAAAAAGCATTCGGAACTGCGACTG	240
Db	181	GGGCTGCAAGACATGCTTCACCCGAGGATTTTAAAAAGCATTCGGAACTGCGACTG	240
Qy	241	AAGTGCATCCAGATAAAATTCGTAGAAATTAAGAAAGACAGAGAGAAAATTCAGGCA	300
Db	241	AAGTGCATCCAGATAAAATTCGTAGAAATTAAGAAAGACAGAGAGAAAATTCAGGCA	300
Qy	301	GTAGCGGAGCATATGAAAGTCTGTGGATCTTAAGAAAGGAGACATCTATGACAAATAT	360
Db	301	GTAGCGGAGCATATGAAAGTCTGTGGATCTTAAGAAAGGAGACATCTATGACAAATAT	360
Qy	361	GGCAAAAGAAAGATTAAATGTGTGNGNGNGNGCTGAGTCAATTTATACGTCATTTGAA	420
Db	361	GGCAAAAGAAAGATTAAATGTGTGNGNGNGNGCTGAGTCAATTTATACGTCATTTGAA	420
Qy	421	TTTGGCTTCACATTCGATMACCAGATGATGCTTTCAGGGAATTTTTGGTGAAGGAC	480
Db	421	TTTGGCTTCACATTCGATMACCAGATGATGCTTTCAGGGAATTTTTGGTGAAGGAC	480
Qy	481	CCATTTTCATTTGATGATCTTGTGAGAACCCCTTTTGAAGACTCTTGTGGGAATGGAAGGGT	540
Db	481	CCATTTTCATTTGATGATCTTGTGAGAACCCCTTTTGAAGACTCTTGTGGGAATGGAAGGGT	540
Qy	541	CCCCGAGAGAGAGAGCCGAGGAGCGGGGCGTTTTCTCTGCTTCAGTGAATTTCCG	600
Db	541	CCCCGAGAGAGAGAGCCGAGGAGCGGGGCGTTTTCTCTGCTTCAGTGAATTTCCG	600
Qy	601	TCCTTTGGAAGTGGATTTTCTCTTTTGAATACAGATTTACTTCATTTGGGTACTAGGT	660
Db	601	TCCTTTGGAAGTGGATTTTCTCTTTTGAATACAGATTTACTTCATTTGGGTACTAGGT	660
Qy	661	CACGGGGGCTTCACATTCGCTTCCACGCACTTGTGTGNGNGCATGGGCAAACTTC	720
Db	661	CACGGGGGCTTCACATTCGCTTCCACGCACTTGTGTGNGNGCATGGGCAAACTTC	720
Qy	721	AAATGATATCACTTCACTAAATGTTTAATGCGAGAAAATCACTCAAAAGAAAT	780
Db	721	AAATGATATCACTTCACTAAATGTTTAATGCGAGAAAATCACTCAAAAGAAAT	780
Qy	781	GTGCGAAGCGGTCAAGAAAGGTGAAGTGAAGAGATGGCAGTTAAAGTCTTAA	840
Db	781	GTGCGAAGCGGTCAAGAAAGGTGAAGTGAAGAGATGGCAGTTAAAGTCTTAA	840
Qy	841	ATTAATGCTGTGCGCGAGAGATGCTTGTGAGAGAGGATGCGAGAGCGACAGAC	900
Db	841	ATTAATGCTGTGCGCGAGAGATGCTTGTGAGAGAGGATGCGAGAGCGACAGAC	900
Qy	901	GTGCTGCGACGCCAGGCTGCGCGGCTCTCCAGCGCCGAGACGCCCGCGCTGCTGTTG	960
Db	901	GTGCTGCGACGCCAGGCTGCGCGGCTCTCCAGCGCCGAGACGCCCGCGCTGCTGTTG	960
Qy	961	CTGAGACAGCGGCTCATTTTCTCTCTAAGAGAGAGGGGACAGACGACGCTTGAGCA	1020
Db	961	CTGAGACAGCGGCTCATTTTCTCTCTAAGAGAGAGGGGACAGACGACGCTTGAGCA	1020
Qy	1021	CCGNGNCTGTGNNCCCCCTCGCTCCNAGCAGGNTTNNAGAAGTNGCAAGAGATG	1080
Db	1021	CCGNGNCTGTGNNCCCCCTCGCTCCNAGCAGGNTTNNAGAAGTNGCAAGAGATG	1080
Qy	1081	NAACGAGAGAGAGAGAGATGTCGAAAGAAAGAAAGTGCACCAAGCAATCACTATAC	1140
Db	1081	NAACGAGAGAGAGAGAGATGTCGAAAGAAAGAAAGTGCACCAAGCAATCACTATAC	1140

QY	1141	CGGACTTAGGCACGCGGTGCATACCCCGACGAGCGCTGGCGGCTCCACCGTGTGGGCAATGGG	1200
Db	1141	CGGACTTAGGCACGCGGTGCATACCCCGACGAGCGCTGGCGGCTCCACCGTGTGGGCAATGGG	1200
QY	1201	TCTGTGCACACCGCTAGGTAGTCACGCTCGGTGACGACTCTCTCGAGGACACACTCGCTCG	1260
Db	1201	TCTGTGCACACCGCTAGGTAGTCACGCTCGGTGACGACTCTCTCGAGGACACACTCGCTCG	1260
QY	1261	GCAGGATTATTCGATTCACGGATCATGTCAGACAGGGTCCAGGAGACGGGGCTACGCGCACG	1320
Db	1261	GCAGGATTATTCGATTCACGGATCATGTCAGACAGGGTCCAGGAGAGCGGGGCTACGCGCACG	1320
QY	1321	GGTGGCGGGG	1330
Db	1321	GGTGGCGGGG	1330

```

US-08-868-288A-2
; Sequence 2, Application US/08668288A
; Patent No. 5922567
; GENERAL INFORMATION:
; APPLICANT: Au-Yang, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868,288A
; FILING DATE: June 3, 1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-4166
; TELEFAX: 415-855-0555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
; US-08-868-288A-2

Query Match 6.9%; Score 91.4; DB 4; Length 1376;
Best Local Similarity 66.5%; Pred. No. 5,5e-18;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

0Y 163 GATTCTATGAAAGTCTAGGCGCTGCAAGACATGCTCACCCGAGAGATTTAAAAAGCA 222
DB 268 GATTTCTTATMAATCTTGGGGGTGCTCGAAGTGCCTTATTAAGATTTAAAAAGGCC 327
0Y 223 TATCGAACAATGGCACTGAAGTGGCATCCAGATAAAAAATCCTGAAGAAATTAAGAAAGCA 282

```

Db 328 TATAGAAACAGCCCTGACCTTCATCCGACGGAACCTG---ATGATCCACAAAGC 384
OY 283 GAGAGAAATTCAGCAAGTAGCGAGGCATATGAAGTGTGCGATGCTAAGAAACGG 342
Db 385 CAGGAGAAATTCAGGATCTGGGTGCTCTTATGAGTTCTGTCAGATAGAGAAACGG 444
OY 343 GACATCTATGACAAATATGSCAAGAAGATTAATGCTGG 383
Db 445 AAACAGTAGATCTTATGTGTGAAGAGATTAAGATGG 485

RESULT 5
US-09-235-373-2
; Sequence 2, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-4166
; TELEFAX: 415-855-0555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
; US-09-235-373-2

Query Match 6.9%; Score 91.4; DB 5; Length 1376;
Best Local Similarity 66.5%; Pred. No. 5.5e-18;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

OY 163 GATTACTATGAAGTCTCTAGGCGTCAGAGACATGCTCACCCGAGATATTAAAAAGCA 222
Db 268 GATTTCATTAAGATCTTGGGGGTGCTCGAAGTGCCTCATTAAGATATTAAAAAGGC 327
OY 223 TATCGGAAACTGCGACCTAGTGCATCCAGATTAATAATCTCGAGATTAAGAAAGCA 282
Db 328 TATAGGAACTAGCCCTGACGCTTCATCCGACCGAAACCTG---ATGATCCACAAGCC 384
OY 283 GAGAGAAATTCAGCAAGTAGCGAGGCATATGAAGTGTGCGATGCTAAGAAACGG 342

Db 385 CAGGAGAAATTCAGGATCTGGGTGCTCTTATGAGTTCTGTCAGATAGAGAAACGG 444
OY 343 GACATCTATGACAAATATGSCAAGAAGATTAATGCTGG 383
Db 445 AAACAGTAGATCTTATGTGTGAAGAGATTAAGATGG 485

RESULT 6
US-09-388-993-2
; Sequence 2, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-4166
; TELEFAX: 415-855-0555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
; US-09-388-993-2

Query Match 6.9%; Score 91.4; DB 5; Length 1376;
Best Local Similarity 66.5%; Pred. No. 5.5e-18;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

OY 163 GATTACTATGAAGTCTCTAGGCGTCAGAGACATGCTCACCCGAGATATTAAAAAGCA 222
Db 268 GATTTCATTAAGATCTTGGGGGTGCTCGAAGTGCCTCATTAAGATATTAAAAAGGC 327
OY 223 TATCGGAAACTGCGACCTAGTGCATCCAGATTAATAATCTCGAGATTAAGAAAGCA 282
Db 328 TATAGGAACTAGCCCTGACGCTTCATCCGACCGAAACCTG---ATGATCCACAAGCC 384
OY 283 GAGAGAAATTCAGCAAGTAGCGAGGCATATGAAGTGTGCGATGCTAAGAAACGG 342
Db 385 CAGGAGAAATTCAGGATCTGGGTGCTCTTATGAGTTCTGTCAGATAGAGAAACGG 444
OY 343 GACATCTATGACAAATATGSCAAGAAGATTAATGCTGG 383
Db 445 AAACAGTAGATCTTATGTGTGAAGAGATTAAGATGG 485


```
RESULT 7
US-08-486-955A-6
; Sequence 6, Application US/08486955A
; Patent No. 5747299
; GENERAL INFORMATION:
; APPLICANT: FATHMAN, Garrison
; APPLICANT: BLOOM, Debra
; TITLE OF INVENTION: Anergy Genes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,955A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A59741-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 672 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-486-955A-6

Query Match 5.7%; Score 75.6; DB 2; Length 672;
Best Local Similarity 58.5%; Pred. No. 2.2e-13;
Matches 148; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 143 CAATCTCGTAAACATGCTGATTAATGATTTAGCGCTGCAGACATGCTCAC 202
DB 23 CACTCTCTACTCCGGGGAATCGTTATACCATGTTCTTGACATGACAAAGAAATGCAACCT 82
QY 203 CCGAGATATTAAAGGATATCGAATATGCGACTGAAGTGCATCCAGATTAATAATC 262
DB 83 CAATATGACATTAAGAAAGTCTATCGGAAGCTGCTGAAGTATCACTGACAAAGAAC 142
QY 263 CTAGAGATTAAGAGAGACAGAGAGAAATTCAGCAAGTAGCGGAGGATATGAATGC 322
DB 143 CTGATTAACCCAGAGCTGCAGAG---CAAGTTTAAGAGATTACACACGACACGCAATCT 199
QY 323 TGTGCGATGTAGAAAGGAGGATATGACAAATATGCGCAAGAGATTAAATGTTG 382
DB 200 TGACACAGCGCCAGAAAGAAACATTATGACAAAGTATGCTGCTGCGTGGCTCTATGTGG 259
QY 383 GNGGNGNGGTGG 395
DB 260 CCGAGCAGTTTGG 272
```

```
RESULT 8
US-08-974-546-2
; Sequence 2, Application US/08974546
; Patent No. 5945287
```

```
GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Puri
; TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastpro for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,546
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0428
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-845-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITTT21
; CLONE: 2525691
; US-08-974-546-2

Query Match 5.5%; Score 73.4; DB 4; Length 2349;
Best Local Similarity 60.8%; Pred. No. 2.2e-12;
Matches 138; Conservative 0; Mismatches 83; Indels 6; Gaps 1;

QY 163 GATTACTATGAAGTTCTAGCGCTGCAGAGACATGCTCACCCGAGATATTAAAGGCA 222
DB 111 GATTATTACAAAGTTCTTGGAATCCCATCGGGGCCAAGAGATGACATCAAGAAAGCC 170
QY 223 TATCGAAGAACTGGCACTGAAGTGCATCCAGATTAATAATCCTGAGAAATTAAGAAAGCA 282
DB 171 TACCGAAGAACTGGCTTGAAGTACCAACCCAGCAAGAAAT-----AAGAACCCAGCCT 224
QY 283 GAGAGAAATTCAGCAAGTAGCGGAGGATATGAAGTGTGCTGCGATGCTTAAGAAAGG 342
DB 225 GAGGAGAACTTAAGAGAGATTGACAGAGGCTATGATGTGCTPAAGTGAACCCCAAGAAAGG 284
QY 343 GACATCTATGACAAATATGCGCAAGAAAGAGATTAAATGTTGNGNGG 389
DB 285 GGCCTGTATGACCAAGTATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 331
```

```
RESULT 9
US-08-472-534-4
; Sequence 4, Application US/08472534
; Patent No. 5919620
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard R
; APPLICANT: Martin, Denis
```

```
? TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM
? TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Neave
? STREET: 1251 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: United States of America
? ZIP: 10020
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/472,534
? FILING DATE:
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: Haley Jr, James F
? REGISTRATION NUMBER: 27,794
? REFERENCE/DOCKET NUMBER: Biovac-2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-596-9000
? TELEFAX: 212-596-9090
? TELEX: 14-8367
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4320 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Streptococcus pneumoniae
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3265..4320
? OTHER INFORMATION: /product= "NH2-terminal portion of
? OTHER INFORMATION: DNA J"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 682..2502
? OTHER INFORMATION: /product= "Heat-Shock Protein 72"
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 682..2502
?
? US-08-472-534-4

Query Match 5.1%; Score 68.2; DB 4; Length 4320;
Best Local Similarity 56.6%; Pred. No. 1.2e-10;
Matches 146; Conservative 0; Mismatches 106; Indels 6; Gaps 1;
```

```
QY 394 GGAGTCATTTTGACGT 411
DB 3502 GCTGTGTTTCGGCGGT 3519

RESULT 10
US-08-879-260-3
? Sequence 3, Application US/08879260
? Patent No. 5935851
? GENERAL INFORMATION:
? APPLICANT: Murthy, Anita E.
? APPLICANT: Gusella, James F.
? TITLE OF INVENTION: TPR-containing Genes
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
? STREET: 1100 New York Ave, N.W., Suite 600
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/879,260
? FILING DATE: 19JUN1997
? CLASSIFICATION: 435
? PRIORITY APPLICATION DATA:
? APPLICATION NUMBER: 60/020,204
? FILING DATE: 20JUN1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Ludwig, Steven R.
? REGISTRATION NUMBER: 36,203
? REFERENCE/DOCKET NUMBER: 0609,4260001/JNG/SRL
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-371-2540
? TELEFAX: 202-371-2540
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1756 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 27..1478
?
? US-08-879-260-3

Query Match 5.0%; Score 66.4; DB 4; Length 1756;
Best Local Similarity 61.2%; Pred. No. 2.4e-10;
Matches 131; Conservative 0; Mismatches 71; Indels 12; Gaps 1;
```

```
RESULT 11
US-08-221-816B-1
; Sequence 1, Application US/08221816B
; Patent No. 5738985
GENERAL INFORMATION:
APPLICANT: Miles, Vincent J.
APPLICANT: Mathews, Michael B.
APPLICANT: Katze, Michael G.
APPLICANT: Witherell, Gary.
APPLICANT: Watson, Julia C.
TITLE OF INVENTION: METHOD FOR SELECTIVE INACTIVATION
TITLE OF INVENTION: OF VIRAL REPLICATION
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/221,816B
FILING DATE: 01-Apr-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7960-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-221-816B-1

Query Match 4.5%; Score 60; DB 2; Length 1515;
Best Local Similarity 59.0%; Pred. No. 1.9e-08;
Matches 124; Conservative 0; Mismatches 80; Indels 6; Gaps 1;

QY 150 GTAACAATGGTGTACTATGAGGCTTCTAGCGTGCAGAGACATGCTCACCAGGA 209
DB 1167 GTACACAGACAGAGATTATACAAATCTTGGAGATGAAAAAGAAATGCCAAAAGCAAGA 1226
QY 210 TATTAAAAAGCATATCGGAACTGGCACTGAAGTGGCATCCAGATAA-----AATCC 263
DB 1227 AATCATTTAAAGCATACCGAAATTAACACTGACGTGACGCCACCCAGACACTTCCAGAACGA 1286
QY 264 TGAGATTAAGAAGAGACAGAGAAATTCAAAGTAGCGGAGGACATATGAAGTGGT 323
DB 1287 AGAAGAAAAAGAAAAAGCTGAGAGAAAGTTTCATTTGACATAGCAGCTGCTAAGAAAGTCTT 1346
QY 324 GTCCGATGCTAAGAAAGGGAGACATCTATGA 353
DB 1347 CTCGATCCAGAAATGAGGAAGAAAGTTTGA 1376

RESULT 12
US-08-143-219-26
; Sequence 26, Application US/08143219
; Patent No. 5670330
```

```
GENERAL INFORMATION:
APPLICANT: Sonenberg, Nahum
APPLICANT: Katze, Michael G.
APPLICANT: Roy, Sophie
APPLICANT: Koromilas, Antonis E.
APPLICANT: Barber, Glen N.
TITLE OF INVENTION: TUMOR-CELL ASSAY METHOD AND KIT
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/143,219
FILING DATE: October 25, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 08/141,244
FILING DATE: October 22, 1993
APPLICATION NUMBER: 07/953,681
FILING DATE: September 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Douglas E. Olson
REGISTRATION NUMBER: 22,798
REFERENCE/DOCKET NUMBER: 204/139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1687 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: BOVINE P58 GENE, FIGURE 10A
US-08-143-219-26

Query Match 4.5%; Score 60; DB 1; Length 1687;
Best Local Similarity 59.0%; Pred. No. 2.1e-08;
Matches 124; Conservative 0; Mismatches 80; Indels 6; Gaps 1;

QY 150 GTAACAATGGTGTACTATGAGGCTTCTAGCGTGCAGAGACATGCTCACCAGGA 209
DB 1282 GTACACAGACAGAGATTATACAAATCTTGGAGATGAAAAAGAAATGCCAAAAGCAAGA 1341
QY 210 TATTAAAAAGCATATCGGAACTGGCACTGAAGTGGCATCCAGATAA-----AATCC 263
DB 1342 AATCATTTAAAGCATACCGAAATTAACACTGACGTGACGCCACCCAGACACTTCCAGAACGA 1401
QY 264 TGAGATTAAGAAGAGACAGAGAAATTCAGCAATGACGAGAGGATATGAAGTCT 323
DB 1402 AGAAGAAAAAGAAAAAGCTGAGAGAAAGTTTCATTTGACATAGCAGCTGCTAAGAAAGTCTT 1461
QY 324 GTCCGATGCTAAGAAAGGGAGACATCTATGA 353
DB 1462 CTCGATCCAGAAATGAGGAAGAAAGTTTGA 1491
```

RESULT 13
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BERT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT9pt-F1s
; US-08-232-463-14

Query Match 4.2%; Score 55.8; DB 1; Length 7218;
Best Local Similarity 3.5%; Pred. No. 9.5e-07;
Matches 9; Conservative 162; Mismatches 87; Indels 0; Gaps 0;

141 AACATCTGCTAAACATGCTGATTAAGTCTAGCGCTGACAGACATGCTC 200
1447 AACAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1388
201 ACCCGAGATATTAAGGATATGGAAGTGGACAGTGAAGTGCATTAATAA 260
1387 RRR 1328
261 TCCTGCAATTAAGAGACAGACAGAAAAATTCAGCAAGTAGCGAGCATTA 320
1327 RRR 1268
321 GCGTGGGATGCTAAGAAACGACATCATGCAATATAGCAAGCAAGATTA 380
1267 RRR 1208
381 TGGNGGNGGNGTGAAG 398
1207 RRR 1190

RESULT 14
US-08-897-340-4
; Sequence 4, Application US/08897340
; Patent No. 5955306
; GENERAL INFORMATION:
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,340
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/715,032
; FILING DATE: 17-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silveri, Jean M.
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: MNI-005CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-897-340-4

Query Match 4.2%; Score 55.2; DB 4; Length 1700;
Best Local Similarity 57.9%; Pred. No. 5.9e-07;
Matches 124; Conservative 0; Mismatches 78; Indels 12; Gaps 1;

154 AACATGCTGATTAAGTCTGAGTCTAGCGCTGACAGACATGCTCAGATAT 213
983 AAGAGGAAGATTTACTACAGATCTCGGAGCTGACAAAGATGCTCGAGAG 1042
214 AAAAAAGCATTCGAAACCTGCGACATGAAGTGCATCCAGTAAAAA-----T 261
1043 AAGAAAGCTTACCGGAAACGGGCTTGATGCACATCCAGATCGGACAGTGG 1102
262 CCGTAGAATTAAGAGAAACGAGAGAAATTCAGCAAGTACGGGAGGATTA 321
1103 GCCGAAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1162
322 CTGTCGATCTAAGAAAGGAGACATCTATGACA 355
1163 CTCTCTATCCCAAGAAAAAGACTCTTATGACA 1196

RESULT 15
US-08-686-417-2
; Sequence 2, Application US/08686417
; Patent No. 5850018
; GENERAL INFORMATION:

```

1  APPLICANT:  Basczynski, Chris
2  APPLICANT:  Barbour, Eric
3  APPLICANT:  Rosowitz, Jeanninge
4  APPLICANT:  Rosichan, Jeffrey L.
5  TITLE OF INVENTION:  AN EXPRESSION CONTROL SEQUENCE FOR
6  TITLE OF INVENTION:  GENERAL AND EFFECTIVE EXPRESSION OF GENES IN PLANTS
7  NUMBER OF SEQUENCES:  5
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE:  MORRISON & FOERSTER
10 STREET:  2000 Pennsylvania Avenue, NW
11 CITY:  Washington
12 STATE:  DC
13 COUNTRY:  USA
14 ZIP:  20006-1888
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE:  Floppy disk
17 COMPUTER:  IBM PC compatible
18 OPERATING SYSTEM:  PC-DOS/MS-DOS
19 SOFTWARE:  PatentIn Release #1.0, Version #1.30
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER:  US/08/686,417
22 FILING DATE:  26-JUL-1996
23 CLASSIFICATION:  800
24 ATTORNEY/AGENT INFORMATION:
25 NAME:  Murashige, Kate H.
26 REGISTRATION NUMBER:  29,959
27 REFERENCE/DOCKET NUMBER:  27112-20038..00
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE:  (202) 887-1500
30 TELEFAX:  (202) 822-0168
31 INFORMATION FOR SEQ ID NO:  2:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH:  3748 base pairs
34 TYPE:  nucleic acid
35 STRANDEDNESS:  double
36 TOPOLOGY:  linear
37 MOLECULE TYPE:  DNA (genomic)
38 FEATURE:
39 NAME/KEY:  CDS
40 LOCATION:  161n(813..962, 2120..2284, 2376..2519, 2605..2880,
41 LOCATION:  2970..3167, 3250..3573)
42 US-08-686-417-2

```


CC
computer

computer read

computer readable medium.

AC T11788;
 DE 07-APR-1996 (first entry)
 KW Mouse cysteine string protein gene.
 KW Mouse; cysteine string protein; anergy; T-lymphocyte;
 KW differential display; cDNA; reverse transcription; probe;
 KW polymerase chain reaction; cloning; gel electrophoresis;
 KW antibody; diagnostic; immunoassay; autoimmune disease;
 KW organ transplantation; vaccine; drug screening; ss.
 OS Mus musculus.
 PN W09600300-A1.
 PD 04-JAN-1996.
 PF 22-JUN-1995; U07958.
 PR 23-JUN-1994; US-265100.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bloom D, Fathman G.
 DR WPI; 96-068884/07.
 DR P-SDB; R90680.
 PT Identifying genes associated with T-cell anergy, and the proteins
 encoded by them - useful to elucidate the anergic state, and
 identify agents associated with induction of anergy
 PS Disclosure; Fig 2; 31pp; English.
 CC The sequence encodes a mouse cysteine string protein isolated from
 CC mouse T-cell clones 11.3.7 and 12.2 by a new procedure for
 CC identification of genes associated with anergy. The protein is
 CC expressed in resting T-cells but not in anergic T-cells. To
 CC isolate proteins of this type, differential display is first
 CC performed by generating cDNA from polyA+ mRNA with an oligo-dT
 CC primer with 2 variable anchor nucleotides at the 3'-end, e.g.
 CC T11786, using reverse-transcriptase. The resulting cDNA
 CC populations are amplified by the polymerase chain reaction,
 CC using the same oligo-dT primer and a 5'-primer which includes
 CC common 4-5 nucleotide sequence combinations, e.g. T11785. After
 CC gel electrophoresis of the products, bands unique to anergised or
 CC non-anergised cells are excised, eluted, amplified, cloned and
 CC sequenced. Genes and proteins (and corresponding antibodies)
 CC arising from this procedure may be used e.g. in monitoring
 CC autoimmune disease, organ transplantation, vaccine
 CC effectiveness, and in drug screening.
 SQ Sequence 672 BP; 171 A; 173 C; 181 G; 147 T;

Query Match 5.7%; Score 75.6; DB 1; Length 672;
 Best Local Similarity 58.5%; Pred. No. 5.3e-12;
 Matches 148; Conservative 0; Mismatches 102; Indels 3; Gaps 1;
 QY 143 CAATCTCGTAAACATGCTGATTACTATGAAGTCTAGGCGTCGACAGACATGCTCAC 202
 DB 23 CACTCTCTACTTCGCGGAATCGTTATACATGTTCTTGACTGACAAAGATGCAACT 82
 QY 203 CCGAGATATTAAAAAGGCATATCGGAACCTGCACTGAAGTGGCATCCAGTAAAAATC 262
 DB 83 CAGATGACATTAAAAAGTCTATCGGAAGCTGCCCTGAAGTATCACCTTACAAGAAC 142
 QY 263 CTGAGATTAAGAGAGACAGAGAAAAATTCAGCAGTAGTGGAGGCATATGAAGTGC 322
 DB 143 CTGATTAACCCAGAGGCTGCAGAAA---CAGTTTAAAGGACATTAAACACCAACGCCATCT 199
 QY 323 TGTGGATGCTAAGAACGAGCATCTATGACAAATATGCAAAAGAGATTAAATGGTG 382
 DB 200 TGACAGACGCCACGAAAGAAACATTATATGACAAAGTATGCGTGGGGGCTATATGG 259
 QY 383 GNGNGGNGGTGG 395
 DB 260 CGGAGCAGTTTGG 272

RESULT 8
 T11789
 ID T11789 standard; cDNA; 672 BP.
 AC T11789;
 DE 07-APR-1996 (first entry)
 KW Human cysteine string protein gene.
 KW Human; cysteine string protein; anergy; T-lymphocyte;

KW differential display; cDNA; reverse transcription; probe;
 KW polymerase chain reaction; cloning; gel electrophoresis;
 KW antibody; diagnostic; immunoassay; autoimmune disease;
 KW organ transplantation; vaccine; drug screening; ss.
 OS Homo sapiens.
 PN W09600300-A1.
 PD 04-JAN-1996.
 PF 22-JUN-1995; U07958.
 PR 23-JUN-1994; US-265100.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bloom D, Fathman G.
 DR WPI; 96-068884/07.
 DR P-SDB; R90680.
 PT Identifying genes associated with T-cell anergy, and the proteins
 encoded by them - useful to elucidate the anergic state, and
 identify agents associated with induction of anergy
 PS Disclosure; Page 19-20; 31pp; English.
 CC The sequence encodes a human cysteine string protein, the absence
 CC of which is associated with anergy in T-cells. Genes with at least
 CC 80% homology to this sequence (e.g. the mouse sequence T11786) may
 CC be used as probes for detection of the anergic state, and may be
 CC isolated by a new procedure. Differential display is first
 CC performed by generating cDNA from polyA+ mRNA with an oligo-dT
 CC primer with 2 variable anchor nucleotides at the 3'-end, e.g.
 CC T11786, using reverse-transcriptase. The resulting cDNA
 CC populations are amplified by the polymerase chain reaction,
 CC using the same oligo-dT primer and a 5'-primer which includes
 CC common 4-5 nucleotide sequence combinations, e.g. T11785. After
 CC gel electrophoresis of the products, bands unique to anergised or
 CC non-anergised cells are excised, eluted, amplified, cloned and
 CC sequenced. Genes and proteins (and corresponding antibodies)
 CC arising from this procedure may be used e.g. in monitoring
 CC autoimmune disease, organ transplantation, vaccine
 CC effectiveness, and in drug screening.
 SQ Sequence 672 BP; 171 A; 173 C; 181 G; 147 T;

Query Match 5.7%; Score 75.6; DB 1; Length 672;
 Best Local Similarity 58.5%; Pred. No. 5.3e-12;
 Matches 148; Conservative 0; Mismatches 102; Indels 3; Gaps 1;
 QY 143 CAATCTCGTAAACATGCTGATTACTATGAAGTCTAGGCGTCGACAGACATGCTCAC 202
 DB 23 CACTCTCTACTTCGCGGAATCGTTATACATGTTCTTGACTGACAAAGATGCAACT 82
 QY 203 CCGAGATATTAAAAAGGCATATCGGAACCTGCACTGAAGTGGCATCCAGTAAAAATC 262
 DB 83 CAGATGACATTAAAAAGTCTATCGGAAGCTGCCCTGAAGTATCACCTTACAAGAAC 142
 QY 263 CTGAGATTAAGAGAGACAGAGAAAAATTCAGCAGTAGTGGAGGCATATGAAGTGC 322
 DB 143 CTGATTAACCCAGAGGCTGCAGAAA---CAGTTTAAAGGACATTAAACACCAACGCCATCT 199
 QY 323 TGTGGATGCTAAGAACGAGCATCTATGACAAATATGCAAAAGAGATTAAATGGTG 382
 DB 200 TGACAGACGCCACGAAAGAAACATTATATGACAAAGTATGCGTGGGGGCTATATGG 259
 QY 383 GNGNGGNGGTGG 395
 DB 260 CGGAGCAGTTTGG 272

RESULT 9
 X20248_03
 Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
 WP Fragment Name Begin End
 WP X20248_00 110000
 WP X20248_01 100001 210000
 WP X20248_02 200001 310000
 WP X20248_03 300001 410000
 WP X20248_04 400001 510000
 WP X20248_05 500001 610000
 WP X20248_06 600001 710000


```

OY 274 GAAAGACACAGAGAAAAATCAAGCAAGTACGGAGGCATTAAGAGTGTGTGCGGATGCT 333
Db 3385 ---GCTGCTAGGAGCAAGTACAAGGAAGTTCAAGAAAGCCTTAAGAACCTTTAGTGTACAGC 34411
OY 334 AAGAAACGGAGACATCTATGACAAATATGGCCAAAGAAAGATTAATAGTGGNAGNGNGNGCT 393
Db 3442 CAAGAAAGCTGCTGCTATGACACAGTATAGTGCTGACGACGCCAAATGAGTGTGTTGTGNGGA 35011
OY 394 GGAAGTCATTTGACAGT 411
Db 3502 GCTGCTGTTGCGCGGT 3519

```

Result	12	Result
V52372		
ID	V52372 standard; DNA: 7174 BP.	
AC	V52372;	
DC	23-OCT-1998 (first entry)	
DE	Streptococcus pneumoniae genome fragment SEQ ID NO:189.	
KW	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;	
KM	computer readable medium; vaccine; pharmaceutical composition; ds.	
OS	Streptococcus pneumoniae.	
PN	W09816931-A2.	
PD	07-MAY-1998.	
PF	30-OCT-1997; U19588.	
PR	31-OCT-1996; US-029960.	
RA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,	
PI	Kunsch CA, Rosen CA;	
DR	WPI; 98-27225/24.	
PT	Computer-readable medium with recorded Streptococcus pneumoniae	
PT	polynucleotide sequences - useful in diagnostic kits and assays, and	
PT	pharmaceutical compositions and vaccines for Streptococcus	
PT	pneumoniae	
PS	Claim 1; Page 1144-1149; 1409pp; English.	
CC	The present invention describes a computer readable medium which has	
CC	the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded	
CC	on it, or a representative fragment or a sequence at least 95% identical	
CC	to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1	
CC	to 391 (V52134 to V52524) are genomic fragments from Streptococcus	
CC	pneumoniae. The present invention also describes an isolated nucleic acid	
CC	molecule encoding a homologue of any of the fragments of the S.pneumoniae	
CC	genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced	
CC	by a process comprising: (a) screening a genomic DNA library using as a	
CC	probe a target sequence defined by any of the sequences in SEQ ID NO:1	
CC	to 391, identifying members of the library which contain sequences	
CC	that hybridize to the target sequence and isolating the nucleic acid	
CC	molecules from the members; or (b) isolating mRNA, DNA or cDNA produced	
CC	from an organism, amplifying nucleic acid molecules whose nucleotide	
CC	sequence is homologous to amplification primers derived from the	
CC	fragment of the S. pneumoniae genome to prime the amplification and	
CC	isolating the amplified sequences. The computer readable medium can be	
CC	used in a computer-based system for identifying fragments of the	
CC	S. pneumoniae genome of commercial importance, or expression modulating	
CC	fragments of the S. pneumoniae genome. Products from the present	
CC	invention can be used in diagnosis kits and assays, and pharmaceutical	
CC	compositions and vaccines for S. pneumoniae.	
SO	Sequence 7174 BP; 1990 A; 1347 C; 1611 G; 2226 T;	
Query Match	5.1%; Score 68.2; DB 1; Length 7174;	
Best Local Similarity	56.6%; Pred. No.2.4e-09;	
Matches 146;	Conservative 0; Mismatches 106; Indels 6; Gaps 1	
OY	154 AACATGGATTTACTATGAAAGTCTAGGCGTGCACAGACATGCTCACCCGAGATATT 213	
Db	1021 AACCACTACTGAATTTATGATGCTGTGGGGGTATCCAAAAAGCCTTCGGCAGCGAAATC 1080	
OY	214 AAAAAGCATATCGGAACAACTGCGACTGAGATGGATCCGATAAAAATCTCGATATAA 273	
Db	1081 AAAAAGGCTTATCGTAAAGCTTTTCAAAAAATATACCCAGATATCAACAAGAGGCT--- 1138	

OY	274	GAAACACAGAGAGAAAATTCAGGCAAGTATGCGAGCATATGAAGTGTCTGCGAGTCT	333
Db	1138	---GGTGTGAGGACAAGTACAGGAAGTTCAAGAACCTATGAGACTTTGAGTACGAC	1194
OY	334	AGAAGACGGGACATCTATGACAATATATGGCAAGAGATTAATAGTGTGNGGNGGNGT	393
Db	1195	CAAAACAGCTGTGCTATAGACAGTATAGTGTCTGACGGCGCAATAGTGTTTGGTGGAA	1254
OY	394	GGAAGTCATTTTGACAGT	411
Db	1255	GCTGTGTGTTTCGGCGGT	1272

Query Match	5.1%	Score 67.2	DB 1	Length 331
Best Local Similarity	59.4%	Pred. No. 8.9e-10		
Matches 133	Conservative 0	Mismatches 85	Indels 6	Gaps 1
DB 166	TACTATGGAAGTCTGAGCGTGACAGACATGCTCACC	CGAGATATTATTAAGGACATAT	225	
DB 107	TACTATGATATCTTATGAGTGTGCGCAAAATTCG	CATCAGAGCCCAATATCAGAGGCTTT	166	
DB 226	CGGAACACTGGCAGTGAAGTGGCATCCAGATATAA	TCTGAGATATAAGAGAGACAGCAG	285	
DB 167	CACAGATGGCGCATGAAGTACACCTGACAAAAT	-----AAGAGCCCAATATGCTGAA	220	
DB 286	AGAAATTCGACGAAGTACGCGGAGCATATGACAT	GAGTCTGCGATGCTTAAGAAAGCGGAC	345	
DB 221	GCAAAATTCAGAGATTCGACAGAACATATGAAAC	ACTCTCAGATGCTTAATAGACGAA	280	
DB 346	ATCTATGACAAATATGCGAAAGAGATTAATATG	TGTGGGAGCG 389		
DB 281	GAGTATGATACCTTGAGACACAGTCTCTTACTAT	GATGTAACGG 324		

```
RESULT 14
T58840_0
WP Sequence split into 6 fragments LOCUS T58840 Accession T58840
WP Fragment Name Begin End
WP T58840_0 1 110000
WP T58840_1 100001 210000
WP T58840_2 200001 310000
WP T58840_3 300001 410000
WP T58840_4 400001 510000
WP T58840_5 500001 580073
ID T58840 standard; DNA; 580073 BP.
AC T58840;
DE 27-MAR-1997 (first entry)
DT Mycoplasma genitalium genome.
KM M. genitalium; DNA; DNA gyrase; origin of replication;
megabase shotgun sequencing method; open reading frame; ORF; ss.
OS Mycoplasma genitalium.
FH Location/Qualifiers
FT cds
FT /*tag= a
FT /label= MG006
FT /note= "Previously identified as MORF-20076, the
FT encoded protein shows 27.59 percentage
FT identity to thymidylate kinase (CDC8)
FT from Saccharomyces cerevisiae"
FT 11252..12040
FT /*tag= b
FT /label= MG009
FT /note= "Previously identified as MORF-20078, the
FT encoded protein shows 35.43 percentage
FT identity to the Bacillus subtilis hypothetical
FT protein covered in accession number
FT GB:D26185.102"
FT 12069..12725
FT /*tag= c
FT /label= MG010
FT /note= "Previously identified as MORF-20079, the
FT encoded protein shows 25.73 percentage
FT identity to DNA primase (dnaE) from
FT Clostridium acetobutylicum"
FT complement (13570..14247)
FT /*tag= d
FT /label= MG012
FT /note= "Previously identified as MORF-20080, the
FT encoded protein shows 31.50 percentage
FT identity to the ribosomal protein S6
FT modification protein (rimK) from Escherichia
FT coli"
FT complement (14396..15217)
FT /*tag= e
FT /label= MG013
FT /note= "Previously identified as MORF-19823, MORF-20080
FT and MORF-20081, the encoded protein shows 33.04
FT percentage identity to 5,10-methylene-tetra-
FT hydrofolate dehydrogenase (tdld) from E. coli"
FT 17474..19243
FT /*tag= f
FT /label= MG015
FT /note= "Previously identified as MORF-20084, the
FT encoded protein shows 32.23 percentage
FT identity to transport ATP-binding protein
FT (msbA) from E. coli"
FT 26478..27344
FT /*tag= g
FT /label= MG023
FT /note= "Previously identified as MORF-20092, the
FT encoded protein shows 45.96 percentage
FT identity to fructose-bisphosphate aldolase
FT (tsr) from B. subtilis"
FT 27345..28448
FT /*tag= h
FT /label= MG024
FT cds
```

```
FT /note= "Previously identified as MORF-19826 and
FT MORF-20093, the encoded protein shows 46.84
FT percentage identity to GTP-binding protein
FT from E. coli"
FT 36987..38978
FT /*tag= i
FT /label= MG032
FT /note= "Previously identified as MORF-20099, the
FT encoded protein shows 26.82 percentage
FT identity to ATP-dependent nuclease (addA)
FT from B. subtilis"
FT 39242..39904
FT /*tag= j
FT /label= MG033
FT /note= "Previously identified as MORF-20100, the
FT encoded protein shows 35.90 percentage
FT identity to glycerol uptake facilitator
FT (g1pf) from B. subtilis"
FT complement (39873..40514)
FT /*tag= k
FT /label= MG034
FT /note= "Previously identified as MORF-20101, the
FT encoded protein shows 48.13 percentage
FT identity to thymidylate kinase (tdk)
FT from B. subtilis"
FT 40543..41787
FT /*tag= l
FT /label= MG035
FT /note= "Previously identified as MORF-20102, the
FT encoded protein shows 30.71 percentage
FT identity to histidyl-tRNA synthetase (h1ss)
FT from Mycobacterium leprae"
FT complement (44751..46277)
FT /*tag= m
FT /label= MG038
FT /note= "Previously identified as MORF-20105, the
FT encoded protein shows 46.83 percentage
FT identity to glycerol kinase (g1pk)
FT from E. coli"
FT complement (46268..47422)
FT /*tag= n
FT /label= MG039
FT /note= "Previously identified as MORF-19831 and
FT MORF-20106, the encoded protein shows 43.20
FT percentage identity to glyceral-3-phosphate
FT dehydrogenase (gud2) from S. cerevisiae"
FT 49377..49643
FT /*tag= o
FT /label= MG041
FT /note= "The encoded protein shows 48.86 percentage
FT identity to phosphohistidinoprotein-hexose
FT phosphotransferase (ptsH) from Mycoplasma
FT capricolum"
FT 50060..51520
FT /*tag= p
FT /label= MG042
FT /note= "Previously identified as MORF-19832 and
FT MORF-20108, the encoded protein shows 41.92
FT percentage identity to spermidine/
FT putrescine transport ATP-binding protein
FT (potA) from E. coli"
FT 51525..52382
FT /*tag= q
FT /label= MG043
FT /note= "Previously identified as MORF-20110, the
FT encoded protein shows 26.51 percentage
FT identity to spermidine/putrescine transport
FT system permease protein (potB) from E. coli"
FT 52366..53220
FT /*tag= r
FT /label= MG044
FT /note= "Previously identified as MORF-20111, the
FT encoded protein shows 29.45 percentage
```

[illegible]

OY	330	TGCTAAGAAACGGGCATTCATGACAATAATGCAAGAAGATTAAATGTGCGNGCG	389
Dd	22565	TGAAGAAAAACCTTAAGCTTTATGACCAGTITTGCTCATGGAAGGTTAAATGCTTCGTGTT	22644
OY	390	NCGTGCA 396 	
Dd	22645	TCATGTA 22651	
RESULT	15		
ID	X40691		
AC	X40691 standard; cDNA; 336 BP.		
DT	18-JUN-1999 (first entry)		
DE	Human secreted protein 5' EST SEQ ID NO: 291.		
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;		
KW	forensic; gene therapy; chromosome mapping; signal peptide; prostate;		
KW	upstream regulatory sequence; cytokine activity; cell proliferation;		
KW	differentiation; haematopoiesis regulation; tissue growth regulation;		
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;		
OS	thrombolytic; anti-inflammatory; tumour inhibition; ds.		
PN	Homo sapiens.		
PF	WO9006550-A2.		
PR	11-FEB-1999.		
PR	31-JUL-1998; IB1232.		
PR	01-AUG-1997; OS-905144.		
PA	(GENSET) GENSET.		
DR	Duchert A, Dumas Milne Edwards J, Lacroix B;		
DR	WPI; 99-153780/13.		
PT	P-PDB; Y11969.		
PR	New isolated prostate-derived nucleic acids - used to develop		
PT	products which may have cytokine, immune regulatory, haematopoiesis		
PR	regulating, anti-inflammatory or tumour inhibition activity		
PS	Claim 1; Page 455-456; 675pp; English.		
CC	X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human		
CC	secreted proteins expressed in prostate, and encode the proteins given		
CC	in Y1176 to Y1193 respectively. The proteins given represent the signal		
CC	peptide and an N-terminal fragment of a secreted protein. The nucleic		
CC	acid sequences can be used for producing secreted human gene products.		
CC	They can also be used to develop products for diagnosis and therapy. The		
CC	proteins obtained may have cytokine activity, cell proliferation and		
CC	differentiation activity, haematopoiesis regulating activity, tissue		
CC	growth regulating activity, reproductive hormone regulating activity,		
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,		
CC	receptor/ligand activity, anti-inflammatory activity, tumour inhibition		
CC	activity or other activities. The products can be used in forensic, gene		
CC	therapy and chromosome mapping procedures. The sequences can also be used		
CC	for obtaining corresponding promoter sequences. The nucleic acids		
CC	encoding the signal peptides can be used for directing extracellular		
CC	secretion of a polypeptide or the insertion of a polypeptide into a		
CC	membrane, or importing a polypeptide into a cell.		
Sequence	336 BP; 92 A; 77 C; 105 G; 60 T;		
Query Match	4.9%; Score 65.6; DB 1; Length 336;		
Best Local Similarity	61.6%; Pred. No. 2.6e-09;		
Matches 141; Conservative 1; Mismatches 74; Indels 13; Gaps 2			
OY	164	ATTACTATGAA GTT AGG CTG CAG ACAC ATC CTC CAC CGA GAT ATT AAAA GG CAT	223
Dd	53	AAT ACT ATG A C AT C C T G G C G T G A A G C C C A G G C G C T C - C C G G A G A G A T C A A G A A G G C C T	111
OY	224	AATCGGAAAATGGCACTGCAAGTGGCANTCCAGATPAAAAATCTTAGATATAAAGAAAGACG	283
Dd	112	ATCGGAAGCTGGCGCTCAAGTACACCCCGACAAGAACCCGGATGGGGCGAC-----	165
OY	284	AAGAGAAAATTCAGCAAGTACGAGGAGCATATGAAAGTCTGTGCGATGTGAAGAAACGGG	343
Dd	165	-----AAGTTTAAATCATATCCCAAGCATATGAAATGCTTTTCAGATCCAAGAAAGG	219
OY	344	ACATCTATGACAAATPRTGGCAAAGAGATTAAATGTTGNGNGNGG	392

Db 220 ATGTTTATGACCAAGCGGAGAGCAGGCAATBVAAGAAGAGGCTCAGG 268

Search completed: June 30, 2000, 14:12:45
Job time: 8232 sec

R;Cneelnam, M.E.; Brion, J.P.; Anderlon, B.H.

Biochem. J. 284, 469-476, 1992
 A>Title: Human homologues of the bacterial heat-shock protein DnaJ are preferentially ex
 A:Reference number: S23508; MID:92287055
 A:Accession: S23508
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1351 <CH>
 A:Cross-references: EMBL:X63368; NID:g32468; PIDN:CAA44968.1; PID:g32469
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:3-69/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 37.0%; Score 635.5; DB 2; Length 351;
 Best Local Similarity 54.4%; Pred. No. 7.8e-44;
 Matches 142; Conservative 30; Mismatches 48; Indels 41; Gaps 9;

QY 1 MVDYEVGVORHASPEDIKAYRKALKWHPDKNPENKEAEKRFQVAEAYVLSDAK 60
 Db 1 MASTYELDVPKASADIRKATRKALQWHPDKNPENKEAEKRFQVAEAYVLSDKH 60
 QY 61 KDIDYKYEKGLNGGGGSHFDSPE---EFGTFPRNPDPVREFPGGRDPSFDFED 117
 Db 61 KREIYDRYEGELGTGTGSRBAAGSGGPFETFRSPREVFREFGSDPFA-ELFDD 119
 QY 118 --PPEDFGRRGRGSRSGTGSFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGF 175
 Db 120 LGPFSEL--ONRG---SRHSGPFETFSFSSFP-----GH---SDF 150
 QY 176 SSTSGGSG--GMGNPKISTSTKMWNGRKITTKRIVNGOERVEVEDGOLSLTINGVAD 234
 Db 151 SSSSFSSPGAGAFRSYSTTTFYQGRITTKRIMNGOERVEVEDGOLSLTINGVAD 210
 QY 235 DDALXERMRGQNVLPAPQA 255
 Db 211 DLARGLELSRREQ-----QPS 226

RESULT 3
 JN0912
 heat-shock protein hsp40 - human
 C:Species: Homo sapiens (man)
 C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 20-Aug-1999
 C:Accession: JN0912; PNM0688; PH0843
 R:Ohtsuka, K.
 Biochem. Biophys. Res. Commun. 197, 235-240, 1993
 A>Title: Cloning of a cDNA for heat-shock protein hsp40, a human homologue of bacterial
 A:Reference number: JN0912; MID:94071949
 A:Accession: JN0912
 A:Molecule type: mRNA
 A:Residues: 1-340 <OHT1>
 A:Cross-references: DDBJ:D49547; DDBJ:DI749; NID:g710654; PIDN:BA08495.1; PID:d1009106
 A:Accession: PNM0688
 A:Molecule type: protein
 A:Residues: 3-49 <OHT2>
 A:Experimental source: Placenta
 R:Hattori, H.; Liu, Y.C.; Tohna, I.; Ueda, M.; Kaneda, T.; Kobayashi, T.; Tanabe, K.; O
 Cell Struct. Funct. 17, 77-86, 1992
 A>Title: Intracellular localization and partial amino acid sequence of a stress-inducibl
 A:Reference number: PH0843; MID:92266438
 A:Accession: PH0843
 A:Molecule type: protein
 A:Residues: 'X', '3-46', 'Y', '48-49 <HAT>
 A:Experimental source: HeLa cell
 C:Genetics:
 A:Gene: GDB:HSPL1
 A:Cross-references: GDB:3929189
 A:Map position: 19p13.2-19p13.2
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 C:Keywords: heat shock; stress-induced protein
 F:4-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match

21.2%; Score 364.5; DB 2; Length 340;

Best Local Similarity 32.9%; Pred. No. 4.8e-22;
 Matches 99; Conservative 46; Mismatches 81; Indels 75; Gaps 13;

QY 3 DVEYEVGVORHASPEDIKAYRKALKWHPDKNPENKEAEKRFQVAEAYVLSDAK 62
 Db 4 DYTQTLGLARGASDEEKRAIKRALKWHPDKNPENKEAEKRFQVAEAYVLSDKR 61
 QY 63 DIYDYKYEKGL-----NGGGGSHFDSPEFETFR-NPDYVREFPGGRDPSFDFED 116
 Db 62 EIDPRYGEGLKSGPGSGGGGANGTS---FSYTFIGDPHAFAEFGGR----- 110
 QY 117 DPEDFGRRGRGSRSGTGSFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 176
 Db 110 NPFDTFFGQRNGEGM-----DIDPFSGF-----MGMG--FT 142
 QY 177 STSGGSGMGNPKISTSTKMWNGRKITTKRIVNGOERVEE--DGLKSLTIN-- 231
 Db 143 NVNFGNS-----RSAQEPARKDDPVT-----HDLRVSLLEIYSGCTKKKISHKRL 190
 QY 231 ----GVADDALXERMRG-----QNVLPAPAGLRPPRPSLRLRXHPCHSKEE 280
 Db 191 NPDGKSRNMDKILITTEYKKGWKEGRKITPFRKGGDQTSNNIPADIVFLKDKRHNIFKR 250
 QY 281 G 281
 Db 251 G 251

RESULT 4
 T09133
 heat shock protein homolog DNJ - Trypanosoma brucei
 N:Alternate names: chaperrone
 C:Species: Trypanosoma brucei
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Sep-1999
 C:Accession: T09133
 R:Bringaud, F.; Vedreone, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays,
 Mol. Biochem. Parasitol. 94, 249-264, 1998
 A>Title: Conserved organization of genes in trypanosomatids.
 A:Reference number: Z16580
 A:Accession: T09133
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-336 <BR1>
 A:Cross-references: EMBL:AF031926; NID:g3452211; PID:g3452212
 A:Experimental source: strain Antat1
 C:Genetics:
 A:Gene: dnj
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 C:Keywords: heat shock; molecular chaperrone; stress-induced protein

Query Match 21.2%; Score 363; DB 2; Length 336;
 Best Local Similarity 50.6%; Pred. No. 6.3e-22;
 Matches 84; Conservative 11; Mismatches 45; Indels 26; Gaps 5;

QY 2 VDVEYEVGVORHASPEDIKAYRKALKWHPDKNPENKEAEKRFQVAEAYVLSDAK 61
 Db 3 IDYTYVGVORHASPEDIKAYRKALKWHPDKNPENKEAEKRFQVAEAYVLSDEK 62
 QY 62 KDIDYKYEKGLNGG-----GGGSHFDSPEFETFRNP--DYVREFPGGRDPP 110
 Db 63 KKIYDSYGEGLKGVDPDSSGPGAGFGFSGGGGTYNFNRDAFKVEESFGSNDPF 122
 QY 111 SPDFEFDPDFGRRGRGSRSGTGSFSAFSGFSGFSGFSGFSGFSGFSGFSGFSG 156
 Db 123 AGG-----DMFGG--GP-----GLHVRFGFGGPHGFGSGFGS 153

RESULT 5
 T04618
 heat shock protein homolog F2009.160 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999

	Query Match	20.9%	Score 358;	DB 21;	Length 338;	
	Best Local Similarity	49.1%;	Pred. No.1.6e-21;			
	Matches	82;	Conservative	20;	Mismatches	39;
					Indels	26;
					Gaps	7.
QY	2 VDYEVVLGVORIASPEDIKKAYRKALKWHPDKNPENKEEARKFEKVQVAEVYLSDAKK	61				
Db	3 IDIYVVLVGRRATSDSIKKATHQLALTKTHPKCKGNRESRSRKRREVSSEADIVLSDENK	62				
QY	62 RDIYDYKKEGILNG-----GG--GGSHFDSPEFEG--FTFRMPD--DVFFEEFGGRD	109				
Db	63 KKIYDYVEBGLKGVPAGSGEGSAGMGTFHGAFPGGVRYRTFSQGDAFNIFPSFGSSDP	122				
QY	110 FSFDFEDPDEDFFGNRRGPFGSRSGRGSGFSFSAASGPFSGSGRSS	156				
Db	123 FAGG-----EEFGGG--GP-----GLHRYVRGGCGPGGFTSGGCS	155				

QY 3 DYEEVLGVQRHASPEDIKKAYRKLALKWHPPDKNPENKEEAERKKQVAEAYEVLSDAKKR 62

Db 4 DYICILGKIGKASDDEIKKAYRKQALKFHPDKN--KSPQAEKFEKFEVAAEVLSDPKKR 61
 QY 63 DIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDBPSPFDFEDPDEF 122
 Db 62 EYIDQFGEGLAGGAGGTGCGGTTRYP-HGDPHATFAAFYXGNSPFXI-----F 111
 QY 123 FGNRRGPPGSRSGRSG---TGSPFSAF---SGFP 148
 Db 112 FGRRMG--GGRDSEMEIDGDPFSAFSGFSMNGYP 143

RESULT 9
 H71526
 Probable heat shock protein 7 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: H71526
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A:Reference number: A71570; MUID:9900809
 A:Accession: H71526
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-392 <ARN>
 A:Cross-references: GB:AE001307; GB:AE001273; NID:93328757; PIDN:AAC67936.1; PID:9332876
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: dnaJ
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:2-67/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 19.2%; Score 330; DB 2; Length 392;
 Best Local Similarity 37.3%; Pred. No. 3.4e-19;
 Matches 107; Conservative 27; Mismatches 89; Indels 64; Gaps 14;

QY 2 VDYEVLGVORHASPEDIKKAYRKALKWHPDKNPKNEKEAEKFKOVAEYVLSDAKK 61
 Db 1 MDYTLIGVATATPPEIKKAYRKALKWHPDKNP-GDAEAEKFEKFEVAAEVLSDPKKR 59
 QY 62 RDIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDBPSPFDFEDPDEF 114
 Db 60 RESIDRKGKGGPPFAGAGGFC-----GAGMGMEALRTFPMKA---FGGDFGNGGCF 108
 QY 115 FEFDF---EDFFGNRRGPPGSRSGRSG---TGSPFSAFSGFPs--FGSGFSSFDGTFTS 163
 Db 109 FEGFLGGLGEAFGMRGSESSESRGASKVHTLSFEAAKGVKEKELVSGYKSGD----- 164
 QY 164 FGSIGHGGLSFSSTSG-----GSGM-----GNFKSISTSTKMY-NGRKITTIRIV 209
 Db 164 -ACGSGC-----ANTAGKGVVCDRCCKSGGVVOSRGFFSNASTCPCDSCGGRVITPCSV 217
 QY 210 ENGGERVEVEDGQLSLTINGVADDAALKEERMR---RGQVNLPA 252
 Db 218 CRGGGRK-----DKRSVHNITAGVDSGMRKMEGIGDAGONGAPRA 259

RESULT 10
 F71623
 Protein with DnaJ domain PFB0090C - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 26-Aug-1999
 C:Accession: F71623
 R:Gardner, M.J.; Tetteilin, H.; Carnucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 116-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743
 A:Accession: F71623
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-354 <GAR>
 A:Cross-references: GB:AE001370; GB:AE001362; NID:93845089; PID:93845091; TIGR:PFB009
 A:Experimental source: clone 307
 C:Genetics:
 A:Gene: PFB0090C
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:21-89/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 19.1%; Score 327; DB 2; Length 354;
 Best Local Similarity 37.6%; Pred. No. 5.3e-19;
 Matches 85; Conservative 23; Mismatches 52; Indels 66; Gaps 8;

QY 3 DYEVLGVORHASPEDIKKAYRKALKWHPDK--NPNKKEAEKFKOVAEYVLSDAK 60
 Db 21 DYIILGVSRDCTIEDIKKAYRKALKWHPDKHNAASKREANMRKRSISEAEVLSDE 80
 QY 61 KRDYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDBPSPFDFEDPFE 120
 Db 81 KRDYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDBPSPFDFEDPFE 126
 QY 121 DFGNRRGPPGSRSGRSGFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGFSG 180
 Db 126 -----YNSPSSPNNGVLFEE-----GSL-FGGSSPFGINP 155
 QY 180 FGGSGMGNFKSISTSTK-----WYNGR---KINTKR 207
 Db 156 RSSGGRYTSKSPSSMDKVEEYVPLVTLLEDYNGTQKILKVTYRKR 201

RESULT 11
 S34630
 dnaJ protein homolog - human
 C:Species: Homo sapiens (man)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S34630; S34631
 R:Chellalish, A.; Davis, A.; Mohanakumar, T. Biochim. Biophys. Acta 1174, 111-113, 1993
 A:Title: Cloning of a unique human homologue of the Escherichia coli DnaJ heat shock
 A:Reference number: S34630; MUID:93326629
 A:Accession: S34630
 A:Molecule type: mRNA
 A:Residues: 1-397 <CHE>
 A:Cross-references: EMBL:L08069; NID:9306713; PIDN:AAC37517.1; PID:9306714
 A:Experimental source: umbilical vein endothelium
 R:Oh, S.; Iwahori, A.; Kato, S. Biochim. Biophys. Acta 1174, 114-116, 1993
 A:Title: Human cDNA encoding DnaJ protein homologue.
 A:Reference number: S34631; MUID:93326630
 A:Accession: S34631
 A:Molecule type: mRNA
 A:Residues: 1-273, 'H', 275-397 <JMA>
 A:Cross-references: EMBL:D13388; NID:9719587; PIDN:BA02656.1; PID:dl003160; PID:g219
 A:Experimental source: fibrosarcoma HT-1080
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:6-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 18.1%; Score 311; DB 2; Length 397;
 Best Local Similarity 32.8%; Pred. No. 1.2e-17;
 Matches 88; Conservative 30; Mismatches 72; Indels 78; Gaps 10;

QY 4 YVEVLGVORHASPEDIKKAYRKALKWHPDKNPKNEKEAEKFKOVAEYVLSDAKKRD 63
 Db 7 YVDVLGVKNATQELKAYRKALKWHPDKNPNGE---KFKQISDAYEVLSDAKRE 62
 QY 64 IYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGG-----RDPSPDFE 115
 Db 63 IYDKGGEQAIKRGAGG-----FGSPMDIFDMFGGGGRMQRERGRKNVYHQL 111
 QY 116 EDPFDFG-----NRGPPRS-----RSRGTGSPFSAFSGFSGS 152
 Db 112 SYTLEDLYNGATRKALKQKNVYCDKSGRGKGAAGVCCPCNCRGTGMQI----- 161

```

QY 153 GFSFDPGFTSFGSLGHGILTSFSSTSRGSGMGNFKSISTKMNKRKIT-TRRIYEN 211
DB 161 -----RIHQIGPEMVQIOISVCEMGCEGHERISPKDRKSCNGRKITREKITLE- 210
QY 212 GOERVEVE---EDGQLSLTINGVADD 236
DB 210 ----VHIDKMKDQ--KITFHGEGDQ 231

RESULT 12
T21991
hypothetical protein F39B2.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21991
R:Dobson, R.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19498
A:Accession: T21991
A>Status: preliminary; translated from GH/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-402 <WIL>
A:Cross-references: EMBL:Z92834; PIDN:CA807390.1; GSPDB:GN00019; CESP:F39B2.10
C:Genetics:
A:Gene: CESP:F39B2.10
A:Map position: 1
A:Introns: 50/3; 202/3; 350/1

Query Match 18.0%; Score 309; DB 2; Length 402;
Best Local Similarity 29.4%; Pred. No. 1.7e-17;
Matches 91; Conservative 38; Mismatches 64; Indels 116; Gaps 11;

QY 4 YEVVLGVORHASPEDIKAYRKALKMHPDKNPKNEKEARKKQVAAVEVLSDAKKRD 63
DB 7 YYDLVGVKRPDASDNELKAYRKALKMHPDKNPDGAEQ----FKQISAYEVLSDPKRQ 62
QY 64 IYDYKKEGLNGGGGSHFDSPEFEGFTFRNPDVFRFFGGNDPFSFDEFEDPFEDFF 123
DB 63 IYDGGGEALQGGGGGGGGRFHNPF-----DVFDMFPG----- 95
QY 124 GNRGPRGRSRGRTGSGFSAFSGSPSGSPTGTSTGSLGHGILT-----SFSST 178
DB 95 GGGGGGGRERERVK-----PTVHNLRVTLDTLTAK-----GVTKKLKISRAT 135
QY 179 SFGSGMGNFKSIISTKMNKRKITTRKRIE-----GVTKKLKISRAT 211
DB 136 CKGCKGLGNEGSAKECSDCKRGKITV-RVIRMGPMVQOMOSHCDSCNGEGSTFLEKDR 194
QY 211 ---NGOERVEVE-----DGQLSLTINGVAD-----DDALKEE 241
DB 195 KKCGKGVKQVKEDEILIEVGITPGMKDGE--KRVFEGKGDEVIGIEKPDGVVVLDEVEHEK 252
QY 242 RMRGQNVL 250
DB 253 FVRKGDML 261

RESULT 13
G72128
heat shock protein J - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: G72128
R:Kalen, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: G72128
A>Status: preliminary
A:Molecule type: DNA

```

```

A:Residues: 1-392 <ARN>
A:Cross-references: GB:AE001588; GB:AE001363; NID:q4376277; PIDN:AD18185.1; PID:q437
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: dnaJ
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match 17.7%; Score 303.5; DB 2; Length 392;
Best Local Similarity 35.5%; Pred. No. 4.6e-17;
Matches 98; Conservative 37; Mismatches 98; Indels 43; Gaps 13;

QY 2 VDYVVLGVORHASPEDIKAYRKALKMHPDKNPKNEKEARKKQVAAVEVLSDAKK 61
DB 1 MDYVILISIKTASAEELIKKALKLVKHPDKN-GAAAEKRRKEVEAEVLSDPQK 59
QY 62 RDIYKKEGLNGGGGSHFDSPEFEGFT--FRNPDVFRFFG--GRDPFSDFPED 117
DB 60 RDSYDRFGKDGPFAGAG-----FGAGAGMGNMEDALRTFMGARGGERGGSPFDG 110
QY 118 PF---EDFFGNRGRGSRGRSGTG-----SFFSAFSGFPS--FGSGFSSFDGTGTFSGS 166
DB 111 LFGGLGEAFGMKRSDPAGARQGSARKVHINLTFEEAAGVEKELVYSGYKSCET-CSQGA 169
QY 167 LGHGLTSFSSSRFGSGM-----GNFKSISTKMN-VNGRKITTRIVENGOERVEVE 220
DB 170 VNPQGIKSCERCK--GSQOVVOSRGFSMASTCPGCGEGRLITDPCSSCRGQGRVK--- 225
QY 221 DGQLSLTINGVADDALKEERM-----RGQNVLPA 252
DB 225 --DKRSVHVHPAGVDSGMRLKMEGYDAGQNGAPS 258

RESULT 14
S42031
LDJ2 protein - leek
C:Species: Allium porrum (leek)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C:Accession: S42031
R:Bessoule, J.J.; Testet, E.; Cassagne, C.
submitted to the EMBL Data Library, February 1994
A:Reference number: S42031
A:Accession: S42031
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-418 <BES>
A:Cross-references: EMBL:X77632; NID:q454913; PIDN:CAA54720.1; PID:q454303
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
P:13-74/Domain: dnaJ amino-terminal homology <DMJ>

Query Match 17.6%; Score 302.5; DB 2; Length 418;
Best Local Similarity 32.7%; Pred. No. 6e-17;
Matches 90; Conservative 25; Mismatches 73; Indels 87; Gaps 11;

QY 4 YEVVLGVORHASPEDIKAYRKALKMHPDK--NPKNEKEARKKQVAAVEVLSDAKK 61
DB 14 YEVVLGVSKNATPEDLKAYRKALKMHPDKGDE-----KKKEIGQAEVVLNDEK 66
QY 62 RDIYKKEGLNGGGGSHFDSPEFEGFTFRNPDVFRFFGGNDPFSFDEFEDPFED 121
DB 67 REIDYQYGEGLKEGMGGGGGVHDPF-----DIFQSFGGG----- 104
QY 122 FFGNRGRGSRGR-----GTGSPFSFAS 145
DB 104 -FGGGSSSRGRGRGDEVDVHPLKVLSDLYNGTSKTLISRNVLCTKCKGSGSGASM 162
QY 146 GPPSP-GSGFSSFDGTGTFSGSLGHGILTSFSSTSRGSGMGNFKSIISTKMNKRKIT 204
DB 163 NCASCGSGMK-----VSIRQIGPMIOQMOPCNCKGTGEMISDRDPOCKGERV 216
QY 205 -TKRIVENGOERVEVE---EDGQLSLTINGVADD 235

```

Db 217 QOKKYLE-----VHEKGMQNGO--KITEPGEADE 244

RESULT 15

S26703

dnaJ protein homolog YDJ1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: dnaJ protein homolog MAS5; protein N2418; protein YNL064c; protein YN

C:Species: Saccharomyces cerevisiae

C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 20-Aug-1999

C:Accession: S26703; A39659; S58714; S62992; S17250

R:Atencio, D.P.; Yaffe, M.P.

Mol. Cell. Biol. 12, 283-291, 1992

A:Title: MAS5, a yeast homolog of dnaJ involved in mitochondrial protein import.

A:Reference number: S26703; MUID:92107179

A:Accession: S26703

A:Molecule type: DNA

A:Residues: 1-409 <ATP>

A:Cross-references: GB:S74758; NID:g241522; PIDN:AAB20771.1; PID:g241523

R:Caplan, A.J.; Douglas, M.G.

J. Cell Biol. 114, 609-621, 1991

A:Title: Characterization of YDJ1: a yeast homologue of the bacterial dnaJ protein.

A:Reference number: A39659; MUID:91332099

A:Accession: A39659

A:Molecule type: DNA

A:Residues: 1-409 <CAP>

A:Cross-references: GB:X56560; NID:g4810; PIDN:CAA39910.1; PID:g4811

R:Bergez, P.; Dolignon, F.; Crouzet, M.

Yeast 11, 967-974, 1995

A:Title: The sequence of a 44 420 bp fragment located on the left arm of chromosome XIV

A:Reference number: S58711; MUID:96021608

A:Accession: S58711

A:Molecule type: DNA

A:Residues: 1-409 <BER>

A:Cross-references: EMBL:U12141; NID:g1314216; PIDN:AAA9647.1; PID:g994823

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994

R:Bergez, P.; Dolignon, F.; Crouzet, M.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S62975

A:Accession: S62975

A:Molecule type: DNA

A:Residues: 1-409 <BEW>

A:Cross-references: EMBL:Z71340; NID:g1301940; PIDN:CAA95937.1; PID:e239683; PID:g130194

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:YDJ1; MAS5

A:Cross-references: SGD:S0005008; MIPS:YNL064c

A:Map position: 14L

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

C:Keywords: heat shock; mitochondrion; stress-induced protein

F:6-70/Domain: dnaJ amino-terminal homology <DNM>

Query Match 17.4%; Score 298.5; DB 2; Length 409;

Best Local Similarity 29.7%; Pred. No. 1,2e-16;

Matches 95; Conservative 36; Mismatches 86; Indels 103; Gaps 14;

QY 4 YVEYLGVQRIASPEDIKAVKALWKHPDKENKKEAEKRFQYAEAYEVLSDAKKRD 63

Db 7 FYDILGVPTATYVEIKAVKALWKHPDKNP--SEAAEKFKKASAAVEILSDPEKRD 64

QY 64 IYDKYGEKGLNGGGGSHDSPEFGFTFRNDDVFRFFGGRDPSFFDFEDFEDFF 123

Db 65 IYDQFGEDGLSGAGAGGFGGFGFC-----DDIFSQFFGA----- 102

QY 124 GNRGRPG--SRSGTGSFFSFGSPFGSGFSSFDGFTGSFSL-----GHGGL 172

Db 102 GGAQRPRGPRGKDIKHEISA-----SLEELYKGRKAKLAKQILCKEGERGK 152

QY 173 TSF--STSGSGSGM-----GNFKSISTSTKMNKRIITTK 206

Db 153 KGAVKKCTSCNGGKIKFVTRQMGPMIORFQTECDVCHGTGDIIDPRCKKSCNGKKVENE 212

QY 207 R-IYENGERVEYVE---EDGOLKSLITNGVADDALKEERRRRCQNVLPAPAGLRPPKP 262

Db 213 RKILE-----VHEPQMKDGO--RIVEKGEAD-----CAPDIVIPGD----- 247

QY 263 PRPASLIRHXPCHLSKEEGE 282

Db 247 -VVEIYSERPHKSPKRDGD 264

Search completed: June 30, 2000, 09:58:20
Job time: 940 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 11:56:22 ; Search time 40.12 Seconds

(Without alignments) 250.502 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716

Sequence: 1 MVDYVEVGVQNRHASPEDIK.....EAERGVEEVEVDQROSLDKT 330

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1230	71.7	241	1	HSJ2_HUMAN
2	1158.5	67.5	242	1	HSJ2_MOUSE
3	847.5	49.4	242	1	HSJ3_MOUSE
4	641.5	37.4	351	1	HS43_HUMAN
5	364.5	21.2	340	1	HS41_HUMAN
6	361	21.0	334	1	DNJ1_DROME
7	311	18.1	397	1	DNJ2_HUMAN
8	305	17.8	397	1	DNJ2_MOUSE
9	304	17.7	369	1	DNJ1_LEPIN
10	302.5	17.6	418	1	DNJ2_ALLPO
11	298.5	17.4	409	1	MMS5_YEAST
12	295	17.2	390	1	DNMJ_MYCPN
13	293	17.1	372	1	DNMJ_BACSU
14	287	16.7	413	1	DNJH_CUCSA
15	286.5	16.7	374	1	DNMJ_CIOAB
16	284	16.6	370	1	DNMJ_ERYRH
17	283.5	16.5	234	1	DNMJ_RHILE
18	281.5	16.4	375	1	DNMJ_BROUY
19	280.5	16.4	389	1	DNMJ_MYCGE
20	280.5	16.3	369	1	DNMJ_NIREU
21	280.5	16.3	417	1	DNJH_ABRNU
22	280	16.3	419	1	DNJH_ABRTH
23	279.5	16.3	378	1	DNMJ_SALTY
24	278	16.2	367	1	DNMJ_COXBU
25	277	16.1	379	1	DNMJ_LACIA
26	276	16.1	352	1	SIS1_YEAST
27	276	16.1	371	1	DNMJ_FRAST
28	274.5	16.0	380	1	DNMJ_BACST
29	271.5	15.8	379	1	DNMJ_LEGPN
30	271	15.8	375	1	DNMJ_ECOLI
31	270.5	15.8	389	1	DNMJ_METMA
32	268.5	15.6	332	1	DNJH_SYNY3
33	266.5	15.5	384	1	DNMJ_RHOCA
34	264.5	15.4	382	1	DNMJ_HAEIN

35	263.5	15.4	377	1	DNMJ_HAEIDU	P48208 haemophilus
36	263	15.3	379	1	DNMJ_STANU	P45555 straphylococ
37	261.5	15.2	511	1	MDJ1_YEAST	P35191 saccharomyc
38	260.5	15.2	364	1	DNMJ_BORBU	P28616 borrelia bu
39	259.5	15.1	352	1	DNMJ_STRPN	P95830 streptococc
40	257.5	15.0	397	1	DNMJ_ALLPO	P03363 allium por
41	256	14.9	404	1	SCJ1_YEAST	P25303 saccharomyc
42	254.5	14.8	355	1	YRVL_YEAST	Q10005 caenorhabdi
43	251.5	14.7	379	1	PSI_CAREL	Q09912 schizosacch
44	246.5	14.4	370	1	PSI_SCHPO	Q9zdy0 rickettsia
45	245.5	14.3	375	1	DNMJ_ACTAC	P77866 actinobacil

ALIGNMENTS

RESULT	ID	Sequence	Score	Match	Length	DB ID	Description
1	HSJ2_HUMAN	US-09-501-714-3	1716	71.7%	241	1	HSJ2_HUMAN
2	HSJ2_MOUSE	US-09-501-714-3	1158.5	67.5%	242	1	HSJ2_MOUSE
3	HSJ3_MOUSE	US-09-501-714-3	847.5	49.4%	242	1	HSJ3_MOUSE
4	HS43_HUMAN	US-09-501-714-3	641.5	37.4%	351	1	HS43_HUMAN
5	HS41_HUMAN	US-09-501-714-3	364.5	21.2%	340	1	HS41_HUMAN
6	DNJ1_DROME	US-09-501-714-3	361	21.0%	334	1	DNJ1_DROME
7	DNJ2_HUMAN	US-09-501-714-3	311	18.1%	397	1	DNJ2_HUMAN
8	DNJ2_MOUSE	US-09-501-714-3	305	17.8%	397	1	DNJ2_MOUSE
9	DNJ1_LEPIN	US-09-501-714-3	304	17.7%	369	1	DNJ1_LEPIN
10	DNJ2_ALLPO	US-09-501-714-3	302.5	17.6%	418	1	DNJ2_ALLPO
11	MMS5_YEAST	US-09-501-714-3	298.5	17.4%	409	1	MMS5_YEAST
12	DNMJ_MYCPN	US-09-501-714-3	295	17.2%	390	1	DNMJ_MYCPN
13	DNMJ_BACSU	US-09-501-714-3	293	17.1%	372	1	DNMJ_BACSU
14	DNJH_CUCSA	US-09-501-714-3	287	16.7%	413	1	DNJH_CUCSA
15	DNMJ_CIOAB	US-09-501-714-3	286.5	16.7%	374	1	DNMJ_CIOAB
16	DNMJ_ERYRH	US-09-501-714-3	284	16.6%	370	1	DNMJ_ERYRH
17	DNMJ_RHILE	US-09-501-714-3	283.5	16.5%	234	1	DNMJ_RHILE
18	DNMJ_BROUY	US-09-501-714-3	281.5	16.4%	375	1	DNMJ_BROUY
19	DNMJ_MYCGE	US-09-501-714-3	280.5	16.4%	389	1	DNMJ_MYCGE
20	DNMJ_NIREU	US-09-501-714-3	280.5	16.3%	369	1	DNMJ_NIREU
21	DNJH_ABRNU	US-09-501-714-3	280.5	16.3%	417	1	DNJH_ABRNU
22	DNJH_ABRTH	US-09-501-714-3	280	16.3%	419	1	DNJH_ABRTH
23	DNMJ_SALTY	US-09-501-714-3	279.5	16.3%	378	1	DNMJ_SALTY
24	DNMJ_COXBU	US-09-501-714-3	278	16.2%	367	1	DNMJ_COXBU
25	DNMJ_LACIA	US-09-501-714-3	277	16.1%	379	1	DNMJ_LACIA
26	SIS1_YEAST	US-09-501-714-3	276	16.1%	352	1	SIS1_YEAST
27	DNMJ_FRAST	US-09-501-714-3	276	16.1%	371	1	DNMJ_FRAST
28	DNMJ_BACST	US-09-501-714-3	274.5	16.0%	380	1	DNMJ_BACST
29	DNMJ_LEGPN	US-09-501-714-3	271.5	15.8%	379	1	DNMJ_LEGPN
30	DNMJ_ECOLI	US-09-501-714-3	271	15.8%	375	1	DNMJ_ECOLI
31	DNMJ_METMA	US-09-501-714-3	270.5	15.8%	389	1	DNMJ_METMA
32	DNJH_SYNY3	US-09-501-714-3	268.5	15.6%	332	1	DNJH_SYNY3
33	DNMJ_RHOCA	US-09-501-714-3	266.5	15.5%	384	1	DNMJ_RHOCA
34	DNMJ_HAEIN	US-09-501-714-3	264.5	15.4%	382	1	DNMJ_HAEIN

Query Match

71.7%

Score 1230;

DB 1;

Length 241;

```
Best Local Similarity 100.0%; Pred. No. 1.2e-87;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
DB 1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
QY 61 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 120
DB 61 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 120
1 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 120
QY 121 DFFGNRRGPRGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 180
DB 121 DFFGNRRGPRGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 180
1 DFFGNRRGPRGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 180
QY 181 GSGGNGNFKSIISTSTKMNNGRKITTKRIYENGOERVEVEDGQLKSLTING 231
DB 181 GSGGNGNFKSIISTSTKMNNGRKITTKRIYENGOERVEVEDGQLKSLTING 231
1 GSGGNGNFKSIISTSTKMNNGRKITTKRIYENGOERVEVEDGQLKSLTING 231

RESULT 2
HSJ2_MOUSE
ID HSJ2_MOUSE STANDARD; PRT; 242 AA.
AC 054946;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DE DNAS PROTEIN HOMOLOG 2 (HEAT SHOCK J2 PROTEIN) (HSJ-2) (MRJ).
GN HSJ2 OR MRJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthelia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-129; TISSUE-EMBRYONIC TROPHBLAST;
RC Hunter P.J., Swanson B.J., Haendel M., Lyons G.E., Cross J.C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS A DNAS-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF035962; AAC16759.1; -
DR HSSP; P25685; 1HDJ.
DR PRAM; PF00226; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1.
KW Chapterone.
FT DOMAIN 1 69 DNAS-LIKE.
SQ SEQUENCE 242 AA; 27012 MW; BAAC2D9EB26F0FA4 CRC64;

Query Match 67.5%; Score 1158.5; DB 1; Length 242;
Best Local Similarity 93.5%; Pred. No. 3.7e-82;
Matches 217; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
DB 1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
QY 61 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 119
DB 61 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 120
1 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 120
QY 120 EDFGNGRRGPRGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 179
DB 121 DFFGNRRGPRGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 180
1 DFFGNRRGPRGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 180
QY 180 FGSGGNGNFKSIISTSTKMNNGRKITTKRIYENGOERVEVEDGQLKSLTING 231
DB 181 FGSGGNGNFKSIISTSTKMNNGRKITTKRIYENGOERVEVEDGQLKSLTING 231
1 FGSGGNGNFKSIISTSTKMNNGRKITTKRIYENGOERVEVEDGQLKSLTING 231

RESULT 3
HSJ3_MOUSE
ID HSJ3_MOUSE STANDARD; PRT; 242 AA.
AC 035723;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DE DNAS PROTEIN HOMOLOG 3 (HEAT SHOCK J3 PROTEIN) (HSJ-3) (MSJ-1).
GN HSJ3 OR MSJ1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthelia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA Berruti G., Perego L., Borgonovo B., Martegani E.;
RC TISSUE-TESTIS;
RX MEDLINE; 98189155.
RT "MSJ-1, a new member of the DNAS family of proteins, is a male germ
RT cell-specific gene product."
RL Exp. Cell Res. 239:430-441(1998).
CC -1- TISSUE SPECIFICITY: TESTIS-SPECIFIC.
CC -1- SIMILARITY: CONTAINS A DNAS-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U95607; AAC13944.1; -
DR HSSP; P25685; 1HDJ.
DR MGD; MG1:1306822; HSJ3.
DR PRAM; PF00226; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1.
KW Chapterone.
FT DOMAIN 1 69 DNAS-LIKE.
SQ SEQUENCE 242 AA; 26678 MW; 5F54F48A0094D84 CRC64;

Query Match 49.4%; Score 847.5; DB 1; Length 242;
Best Local Similarity 70.0%; Pred. No. 2.6e-58;
Matches 163; Conservative 30; Mismatches 37; Indels 3; Gaps 3;

QY 1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
DB 1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
1 MVDYEVLVQVRRASPEDIKKAYRKALKWHPDKNPENKEERKFKQVAEYEVLSAK 60
QY 61 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 118
DB 61 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 120
1 KRDIYDKYKEGLNGGGGSHFDSPEFFGTFRRNPDDVFRFFGGRDPFSFDFEDPF 120
QY 119 FDFGNGRRGPRGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 178
DB 121 LENFGDRKSTGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 179
1 LENFGDRKSTGSRSGTSFSAFSGFSFSGSPTGTFSGSLGHGGLTSSSTS 179
QY 179 SFGGNGNGNFKSIISTSTKMNNGRKITTKRIYENGOERVEVEDGQLKSLTING 231
DB 180 SCGGGAGNGKYSVSTSTELINKKITTKRIYENGOERVEVEDGQLKSLTING 232
1 SCGGGAGNGKYSVSTSTELINKKITTKRIYENGOERVEVEDGQLKSLTING 232

RESULT 4
HS43_HUMAN
ID HS43_HUMAN STANDARD; PRT; 351 AA.
AC P25686;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
```

```

DE HEAT SHOCK 40 KD PROTEIN 3 (DNAJ PROTEIN HOMOLOG 1) (HSP-1).
GN HSPF3 OR HSJ1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE; 92287055.
RA Cheetham M.E., Brion J.-P., Anderton B.H.;
RT "Human homologues of the bacterial heat-shock protein DnaJ are
RT preferentially expressed in neurons.";
RL Biochem. J. 284:469-476(1992).
RN [2]
RP REVISION TO 214.
RA Cheetham M.E.;
RL Submitted (JUL-1998) to the SWISS-PROT data bank.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: HSJ1A AND HSJ1B (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: BRAIN (NEURONAL LAYERS). WEAKLY, IN SKELETAL
CC MUSCLE AND SPLEEN.
CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63368; CAA44968.1; ALT_SEQ.
DR EMBL; X63368; CAA44969.1; ALT_SEQ.
DR PIR; S23508; S23508.
DR HSP; P25685; 1HDJ.
DR MTM; 604139; -.
DR PROSITE; PS00636; DNAJ_1; 1.
DR PROSITE; PS50076; DNAJ_2; 1.
DR PFAM; PF00226; DnaJ; 1.
DR Chapterone; Neurone; Alternative splicing.
KW DOMAIN
FT VARSPIC 275 277 GGR -> DVE (IN ISOFORM HSJ1A).
FT VARSPIC 278 351 MISSING (IN ISOFORM HSJ1A).
FT SEQUENCE 351 AA; 38783 MW; 1FA290C676465F3 CRC64;
SQ

```

```

ID HS41_HUMAN STANDARD; PRT; 340 AA.
AC P25685.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEAT SHOCK 40 KD PROTEIN 1 (HEAT SHOCK PROTEIN 40) (HSP40) (DNAJ
DE PROTEIN HOMOLOG 1) (HDJ-1).
GN HSPF1 OR DNAJ1 OR HDJ1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA.
RX MEDLINE; 92093635.
RA Raabe T., Manley J.L.;
RT "A human homologue of the Escherichia coli DnaJ heat-shock protein.";
RL Nucleic Acids Res. 19:6645-6645(1991).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-48.
RC TISSUE-PLACENTA.
RX MEDLINE; 94071949.
RA Ohtsuka K.;
RT "Cloning of a cDNA for heat-shock protein hsp40, a human homologue of
RT bacterial DnaJ.";
RL Biochem. Biophys. Res. Commun. 197:235-240(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA.
RX MEDLINE; 97131529.
RA Hata M., Okumura K., Seto M., Ohtsuka K.;
RT "Genomic cloning of a human heat shock protein 40 (Hsp40) gene (HSPF1)
RT and its chromosomal localization to 19p13.2.";
RL Genomics 38:446-449(1996).
RN [4]
RP STRUCTURE BY NMR OF 1-76.
RX MEDLINE; 96291433.
RA Qian Y.Q., Patel D., Hartl F.-U., Mccoll D.J.;
RT "Nuclear magnetic resonance solution structure of the human Hsp40
RT (HDJ-1) J-domain.";
RL J. Mol. Biol. 260:224-235(1996).
CC -1- SIMILARITY: INTERACTS WITH HSP70 AND CAN STIMULATE ITS ATPASE
CC ACTIVITY STIMULATES THE ASSOCIATION BETWEEN HSC70 AND HIP.
CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62421; CAA44287.1; -.
DR EMBL; D49547; BA008495.1; -.
DR EMBL; D83429; BA012819.1; -.
DR PIR; S20062; S20062.
DR PDB; 1HDJ; 08-NOV-96.
DR PROSITE; PS00636; DNAJ_1; 1.
DR PROSITE; PS50076; DNAJ_2; 1.
DR PFAM; PF00226; DnaJ; 1.
DR PFAM; PF01556; DnaJ_C; 1.
DR Heat shock; Chapterone; 3D-structure.
KW DOMAIN
FT VARSPIC 1 70 DNAJ-LIKE.
FT VARSPIC 11 11 L -> Q (IN REF. 1).
FT VARSPIC 13 28 RGASDEEIKRAVRQA -> AALRGDQAGLPPG (IN
FT VARSPIC 68 68 REF. 1).
FT VARSPIC 81 136 G -> L (IN REF. 1).
FT VARSPIC 81 136 SGGANGTSFSTFGDPHAFAEFGGRNPFDFGORG
FT VARSPIC 81 136 EEGMDIDPFSGFPM -> TAEPMVPLSATSHMETLMPCL
FT VARSPIC 81 136 LSSVAELPLTFPGSGGGRKAWTLMTHTSLASLW (IN
FT VARSPIC 81 136 REF. 1).
FT VARSPIC 81 136 R -> C (IN REF. 1).
FT CONFLICT 150 150

```


FT CONFLICT 183 183 M -> T (IN REF. 1).
 FT CONFLICT 320 320 V -> A (IN REF. 1).
 SQ SEQUENCE 340 AA; 38044 MW; 17545098B0C196DF CXC64;

Query Match
 Best Local Similarity 21.2%; Score 364.5; DB 1; Length 340;
 Matches 99; Conservative 46; Mismatches 81; Indels 75; Gaps 13;

QY 3 DYEVVLGVORHASPEDIKKAYRKALAKWHPDKNPNKNEKEERKFOVAEYVLSDAKR 62
 DB 4 DYQVQLGLARAGSDEIKRAIRKRLRHPDKNKE--PGAEEKFEIADYVLSDPKR 61
 QY 63 DIYDKYKEGL-----NGGGGGSHFSDPFERGFTR-NPDDVFERFGDRPFESDFE 116
 DB 62 EIFFDYGEGELKSGSGSGGANGTS--FSYTFHDDPHAMFEFFGGR----- 110
 QY 117 DPEDFPFRNRGRPSRSGRSGTGFSSAFSGFSGSGSFDTGFTSGSLGHGLSFS 176
 DB 110 NPFDTFFGQRNGEEGM-----DIDDPFGFP-----MGMGG--FT 142
 QY 177 STSFGSGMGNFKSTSTKMWNGRKITTKRIVENGQERVEVE--DQGLSLTIN---- 231
 DB 143 NVNNGRS-----RSGQEPARKKQDPPTV-----HDLRSLSEETYSCTKKMKISHRL 190
 QY 231 -----GVADDALKEERRRG-----QNVLPAPAGLRPPKPPRPAISLRHXPHCLSKEE 280
 DB 191 NPDKSKINEDKILITEVKKMGKGGTKITFPKGGDQTSNNIPADIVFLKDKPHNIFKRD 250
 QY 281 G 281
 DB 251 G 251

RESULT 6

DNJ1 DROME STANDARD; PRT; 334 AA.

AC 024133;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNAJ PROTEIN HOMOLOG 1 (DROJ1).
 GN DNAJ-1 OR DROJ1.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 NC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lee J.Y., Palter K.B.
 RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U34904; AAC23584.1; -
 DR HSSP: P25685; 1HDJ.
 DR FLYBASE: FBgn0015657; DnaJ-1.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PROSITE: PS50076; DNAJ_2; 1.
 DR PFAM: PF00226; DnaJ; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 KW Chapterone; Heat shock.
 FT DOMAIN 1 70
 SQ SEQUENCE 334 AA; 37100 MW; 6056A235D6F2DC89 CRC64;

Query Match
 Best Local Similarity 21.0%; Score 361; DB 1; Length 334;
 Matches 99; Conservative 37; Mismatches 91; Indels 66; Gaps 9;

QY 3 DYEVVLGVORHASPEDIKKAYRKALAKWHPDKNPNKNEKEERKFOVAEYVLSDAKR 62
 DB 4 DYFIILGLERKASDEIKKAYRKALAKYHPDKN--KSPQAEERKFEIADYVLSDAKR 61
 QY 63 DIYDKYKEGLNGGGGGSHFSDPFERGFTR-NPDDVFERFGDRPFESDFE 121
 DB 62 DIFDNYGEGDLKGGQPPDGGGQGAATYQFHDPDRATFQFGSSDPF----- 111
 QY 122 FFGNRKRPGRSGRSGTGFSSAFSGFSGSGSFSDTGTSGSLGHGLSFSSTSG 181
 DB 111 -----GAF-----TGDDNMFSGGGGNTNIFMNIG 137
 QY 182 GSGMGNFKSTSTKMWNGRKITTKRIVENGQERVEVEEDGQLSLINGVA--DDALX 239
 DB 138 GDDMFAPNAQAPSKKQDPPIEHDLFVS-----LEVDGCTKKKISMATGSGPYK 192
 QY 240 EERRRGQNVLPAPAGLR-----PKPPRP-ASLLRHXPHCLSKEE 281
 DB 193 EEKYLK-ITVKGPKMKAGTKITFPQEGDSAPNKPADIVFLIRKPHSLFKREG 244

RESULT 7

DNJ2_HUMAN STANDARD; PRT; 397 AA.

AC P31689;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE DNAJ PROTEIN HOMOLOG 2 (HSJ-2).
 GN HSJ2 OR DNAJ2 OR HDJ2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 93326630.
 RA Oh S., Iwahori A., Kato S.;
 RT "Human cDNA encoding dnaJ protein homologue.";
 RL Biochim. Biophys. Acta 1174:114-116(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 93326629.
 RA Chelliah A., Davys A., Mohanakumar T.;
 RT "Cloning of a unique human homologue of the Escherichia coli DNAJ
 RT heat shock protein.";
 RL Biochim. Biophys. Acta 1174:111-113(1993).
 CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN AND THE FOUR CXXCGXG
 CC MOTIFS FOUND IN DNAJ PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D13388; BAA02656.1; -
 DR EMBL: L08069; AAC37517.1; -
 DR PIR: S34631; S34631.
 DR HSSP: P25685; 1HDJ.
 DR MIM: 602837; -
 DR PRINTS: PR00625; DNAJPROTEIN.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PROSITE: PS00637; DNAJ_CXXCGXG; 1.
 DR PROSITE: PS50076; DNAJ_2; 1.
 DR PFAM: PF00226; DnaJ; 1.

CC WITH GRPE, THE ATPASE ACTIVITY OF DNAB (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAB, AND TO EUKARYOTIC
 CC DNAB-LIKE PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF007813; AAC35417.1; -
 DR HSSP: U72647; AABI7396.1; -
 DR HSSP: P08622; 1XBL.
 DR PROSITE: PS00636; DNAB_1; 1.
 DR PROSITE: PS00637; DNAB_2; 1.
 DR PROSITE: PS00637; DNAB_CXKXGXG; FALSE_NEG.
 DR PRAM: PF00226; DnaJ; 1.
 KW Chapterone; DNA replication; Heat shock; Repeat.
 FT DOMAIN 3 72 DNAB-LIKE.
 FT REPEAT 150 157 CXKXGXG MOTIF.
 FT REPEAT 167 174 CXKXGXG MOTIF.
 FT REPEAT 189 196 CXKXGXG MOTIF.
 FT REPEAT 203 210 CXKXGXG MOTIF.
 SQ SEQUENCE 369 AA; 39939 MW; 408B5B11457DCDEF CRC64;

Query Match 17.7%; Score 304; DB 1; Length 369;
 Best Local Similarity 35.2%; Pred. No. 2.1e-16;
 Matches 92; Conservative 31; Mismatches 70; Indels 68; Gaps 12;

QY 4 YVEVLVQVRHASPEDIKAYRKALAKWHPDKNPENKEAEKRFQVAEAYVLSDAKRD 63
 DB 6 YVDILGYSKSANDBEIKSAVKRLAIKYPDKNGNK-ESEKREAEVAEAILNDPKRKQ 64
 QY 64 IYDKYKREGLNGGGG---GSHPDSPFEFGFTFRNPDDVREFFGGDPPSFDFEDPFE 120
 DB 65 AYDQFAGVSGAGGAGGCGGAYTDFSDIFG-----DIGDFGGRSSGFG----- 111
 QY 121 DFGNRRGP-RGSRSR-----GTG-SFSPAFSGFP 148
 DB 111 --GGRAGSPQRGSDLRKLNLEVSLEDAALGREYKLEIRLSCVDCNCSGSKSSPPTCP 168
 QY 149 SF-GSGFSSFDGTGFTSGSL-----GHGGLTSPSSSTFGSGGNGNFKSISTSTMVNGRK 202
 DB 169 DCGSGGQIRRTGFFSVATTCPTCRGKGTIISNPCRCGCGGQIQE-KRRTINIKIPPG-- 226
 QY 203 ITRKRIYNGQERVEVEDGQ 223
 DB 226 -----VETG-SRLKVSGE 239

RESULT 10
 DNJ2.ALLPO STANDARD; PRT; 418 AA.
 AC P42824;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNAB PROTEIN HOMOLOG 2.
 GN LDJ2.
 OS Allium porrum (Leek).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;
 OC Alliaceae; Allium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bessoule J.J., Testet E., Cassagne C.;
 RT Cloning of a new isoform of a DnaJ protein from Allium porrum
 RT epidermal cells";
 RL Plant Physiol. Biochem. 32:723-727(1994).
 CC -1- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY

CC IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS A DNAB-LIKE DOMAIN AND THE FOUR CXKXGXG
 CC MOTIFS FOUND IN DNAB PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X77632; CAAS4720.1; -
 DR HSSP: P25685; 1HDI.
 DR PRINTS: PR00625; DNABPROTEIN.
 DR PROSITE: PS00636; DNAB_1; 1.
 DR PROSITE: PS00637; DNAB_CXKXGXG; 1.
 DR PROSITE: PS50076; DNAB_2; 1.
 DR PRAM: PF00226; DnaJ; 1.
 DR PRAM: PF00684; DnaJ_CXKXGXG; 1.
 KW Chapterone; Repeat; Prenylation; Lipoprotein; Multigene family.
 FT DOMAIN 11 76 DNAB-LIKE.
 FT DOMAIN 83 88 POLY-GLY.
 FT DOMAIN 81 108 GLY-RICH.
 FT REPEAT 148 155 CXKXGXG MOTIF.
 FT REPEAT 164 171 CXKXGXG MOTIF.
 FT REPEAT 191 198 CXKXGXG MOTIF.
 FT REPEAT 207 214 CXKXGXG MOTIF.
 FT LIPID 415 415 FARNESYL (BY SIMILARITY).
 SQ SEQUENCE 418 AA; 46584 MW; DCE2A4DF192329E6 CRC64;

Query Match 17.6%; Score 302.5; DB 1; Length 418;
 Best Local Similarity 32.7%; Pred. No. 3.1e-16;
 Matches 90; Conservative 25; Mismatches 73; Indels 87; Gaps 11;

QY 4 YVEVLVQVRHASPEDIKAYRKALAKWHPDK-NPENKEAEKRFQVAEAYVLSDAK 61
 DB 14 YVEVLVSKNATPEDEKKAIRKALAKWHPDKGDEP-----KREIQVAEVLNDPK 66
 QY 62 RDIYDKYKREGLNGGGGSHDPSPEFGFTFRNPDDVREFFGGDPPSFDFEDPFE 121
 DB 67 RETIDYGEGLGKEKGGGGGVHDP-----DIQSPFGGG----- 104
 QY 122 FFGNRRGPGSR-----GTGSPFSAFS 145
 DB 104 -FEGGSSRGRRQRGDEYVHPLKYSLEDLYNGTSKILSIHNVLCITCKCGKSGKASM 162
 QY 146 GFPSF-GSGFSSFDGTGFTSGSLGHGGLTSPSSSTFGSGGNGNFKSISTSTMVNGRKIT 204
 DB 163 NCASCGSGMK-----VSRQLGFMQIQOMQHPNCKEKGTEGEMTSIDKRCPOCKGERV 216
 QY 205 -TRKRIYNGQERVEE---EDGQLKSLTINGVADD 235
 DB 217 QOKKYLE-----VHYEKGQNGQ--KITEPGEADE 244

RESULT 11
 MAS5_YEAST STANDARD; PRT; 409 AA.
 AC P25491;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MITOCHONDRIAL PROTEIN IMPORT PROTEIN MAS5 (PROTEIN YDJ1).
 GN MAS5 OR YDJ1 OR YNL064C OR N2418 OR YNL2418C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE: 92107179.
 RA Atencio D.P., Yaffe M.P.;
 RT "MASS5, a yeast homolog of DnaJ involved in mitochondrial protein
 RT import.";
 RL Mol. Cell. Biol. 12:283-291(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91332099.
 RA Caplan A.J., Douglas M.G.;
 RT "Characterization of YDJ1: a yeast homologue of the bacterial dnaJ
 RT protein.";
 RL J. Cell Biol. 114:609-621(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=5288C / FY1676;
 RX MEDLINE: 96021608.
 RA Berger P., Dolignon F., Crouzet M.;
 RT "The sequence of a 44 420 bp fragment located on the left arm of
 RT chromosome XIV from *Saccharomyces cerevisiae*.";
 RL Yeast 11:967-974(1995).
 RN [4]
 RP ERRATUM.
 RX MEDLINE: 97060022.
 RA Berger P., Dolignon F., Crouzet M.;
 RL Yeast 12:297-297(1996).
 RN [5]
 RP ISOPRENOLD.
 RX MEDLINE: 92406811.
 RA Caplan A.J., Tsai J., Casey P.J., Douglas M.G.;
 RT "Parnesylation of YDJ1p is required for function at elevated growth
 RT temperatures in *Saccharomyces cerevisiae*.";
 RL J. Biol. Chem. 267:18890-18895(1992).
 CC -1- FUNCTION: PROBABLY INVOLVED IN MITOCHONDRIAL PROTEIN IMPORT. IS
 CC ALSO REQUIRED FOR EFFICIENT TRANSLLOCATION OF PRE-PRO-ALPHA-FACTOR.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED IN A PERINUCLEAR RING AS WELL
 CC AS IN THE CYTOPLASM (ACCORDING TO REF. 2).
 CC -1- INDUCTION: MASS5 IS A HEAT SHOCK GENE WHOSE EXPRESSION INCREASES
 CC MODERATELY AT ELEVATED TEMPERATURES.
 CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN AND THE FOUR CXXCGXG
 CC MOTIFS FOUND IN DNAJ PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S74758; AAB20771.1; -;
 DR EMBL: X56560; CAA39910.1; -;
 DR EMBL: U12141; AAA99647.1; -;
 DR EMBL: Z71340; CAA95937.1; -;
 DR PIR: A39659; A39659.
 DR PIR: S26703; S26703.
 DR HSSP: P25685; 1HDJ.
 DR SCD; L0002503; YDJ1.
 DR PRINTS; PR00625; DNAJPROTEIN.
 DR PROSITE; PS00636; DNAJ_1; 1.
 DR PROSITE; PS00637; DNAJ_CXXCGXG; 1.
 DR PROSITE; PS50076; DNAJ_2; 1.
 DR PRAM; PF00226; DnaJ; 1.
 DR PRAM; PF00684; DnaJ_CXXCGXG; 1.
 DR PRAM; PF01556; DnaJ_C; 1.
 KW Chaperone; Heat shock; Transport; Protein transport; Mitochondrion;
 KM Repeat; Prenylation; Lipoprotein.
 FT DOMAIN 4 72 DNAJ-LIKE.
 FT DOMAIN 73 103 GLY-RICH.
 FT REPEAT 143 150 CXXCGXG MOTIF.
 FT REPEAT 159 166 CXXCGXG MOTIF.
 FT REPEAT 185 192 CXXCGXG MOTIF.
 FT REPEAT 201 208 CXXCGXG MOTIF.
 FT LIPID 406 406 FARNESYL.

SQ SEQUENCE 409 AA; 44670 MW; EA539F3618DD9CF2 CRC64;
 Query Match 17.4%; Score 298.5; DB 1; Length 409;
 Best Local Similarity 29.7%; Pred. No. 6.1e-16;
 Matches 95; Conservative 36; Mismatches 86; Indels 103; Gaps 14;
 QY 4 YVEYVGVRHASPEDIKAKYRKALKMHPDKNPEKKEAEKRFVNAVAYVLSAKRD 63
 DB 7 FYDLIGVPTATDVEIKKAYRKALKYHPDKNP--SEEAARKEFASAYEILSPDKRD 64
 QY 64 IYDKYKREGLNGGGGSHFDSPEFGFTFRNDDVFEFGGRDPFSFDFEDFEDFF 123
 DB 65 IYDQGEDEGLSGAGAGGFPFGGFG--DDIFGFFGA----- 102
 QY 124 GNRGPRG-SRSRGTSFSAFSGFSPSGSFSGFTSGSL-----GHGL 172
 DB 102 GGAQPRPQPGKDKIHEISA-----SLEELYKRTAKLALNKQILCKEGRGK 152
 QY 173 TSF--SSTSPGSGM-----GNFKSISTKRVNKRKITK 206
 DB 153 KGAVKCTSCNGGIGIKFYTRQMGPIORFQTECDVCHGTGDIIDPKDCKSCNGKYEVE 212
 QY 207 R-IYENGQERVEE---EDGLKSLITGVADDDALXERRRQGVLPAPAGLRPPKP 262
 DB 213 RKLE-----VHPEGMKDGQ--RIVKGEAD-----QAPDVTPGD----- 247
 QY 263 PRPASLRKXPHCLSKERGE 282
 DB 247 --VVEIVSERPHKSPKRDGD 264
 RESULT 12
 ID DNAS_MYCPN STANDARD; PRT; 390 AA.
 AC P78004;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DNAS PROTEIN.
 GN DNAS.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE: 97105885.
 RA Himmelfreuch R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma*
 RT *pneumoniae*.";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- FUNCTION: INTERACTS WITH DNAJ, TO DISASSEMBLE A PROTEIN COMPLEX
 CC AT THE PHAGE LAMBD ORIGIN OF REPLICATION. STIMULATES, JOINTLY
 CC WITH GREP, THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAS, AND TO EUKARYOTIC
 CC DNAS-LIKE PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000015; AAB95781.1; -;
 DR HSSP; P25685; 1HDJ.
 DR PROSITE; PS00636; DNAS_1; 1.
 DR PROSITE; PS50076; DNAS_2; 1.
 DR PROSITE; PS00637; DNAS_CXXCGXG; 1.
 DR PRAM; PF00226; DnaJ; 1.

```

DR PFAM: PF00684; DnaJ_CXXCXXGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
KW Chaperone; DNA replication; Heat shock; Repeat.
FT DOMAIN 5 79 DnaJ-LIKE
FT REPEAT 165 172 CXXCXXGXG MOTIF.
FT REPEAT 183 190 CXXCXXGXG MOTIF.
FT REPEAT 209 216 CXXCXXGXG MOTIF.
FT REPEAT 223 230 CXXCXXGXG MOTIF.
SQ SEQUENCE 390 AA; 43167 MW; 442E1A6304E8697E CRC64;

Query Match 17.2%; Score 295; DB 1; Length 390;
Best Local Similarity 35.8%; Pred. No. 1.1e-15;
Matches 92; Conservative 24; Mismatches 67; Indels 74; Gaps 13;

QY 3 DYEEVLGVORHASPEDIKAYRKLAKMHPDKNPENKE-----EMERKKQVAEAYEVLSD 58
DB 7 DYEEVLGVSRSATADIRAKRLAMQYHPRHKGEGTYOKONEKREKVEAYEVLSD 66
QY 59 AKKRDIYDKYKEGLNGGGGGSHDFEFERFTFRNPDVFRFFFGGRDPFDF-FED 117
DB 67 TEKRGMYDRFGHEGLNAGSF-----HETGF--NPFDLFNVFG--EGSFDMOGS 113
QY 118 PFEDEFGRRRRGRSRR-----GTG---SFF-----SAF 144
DB 114 PFDLIFS-----RKKRSQNRYVLPYDELAIVGVDISFEWNTGCTRTIETKKRVTSAC 168
QY 145 SGFSPFGSGSFPTGFTSGSL-GHGGLTFSSTSG-----GSGMNEKISTST 195
DB 169 DGFAGEG-----ETGWSCHSCSGNGFILKNQISFTGYVSOMLCOSGGCGKQAKHK 223
QY 196 KAVNGRKI---TTKRI 208
DB 224 KTCGSKYKVPYTKETI 240

RESULT 13
DnaJ_BACSU
ID DnaJ_BACSU STANDARD; PRT; 372 AA.
AC P17631;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DnaJ PROTEIN.
OS
GN DnaJ.
OC Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / MB11;
RX MEDLINE: 92250426.
RA Weizstein M., Voelker U., Dedio J., Loebau S., Zuber U.,
RA Schlussovohl M., Hergel C., Hecker M., Schumann W.;
RT "Cloning, sequencing, and molecular analysis of the dnaK locus from
RT Bacillus subtilis.";
RL Bacteriol. 174:3300-3310(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.;
RT Submitted (May-1996) to the EMBL/GenBank/DBD databases.
RN [3]
RP SEQUENCE OF 1-24 FROM N.A.
RC STRAIN-168 / MB11;
RX MEDLINE: 90245667.
RA Weizstein M., Dedio F., Schumann W.;
RT "Complete nucleotide sequence of the Bacillus subtilis dnaK gene.";
RN [4]
RP SEQUENCE OF 257-372 FROM N.A.
RC STRAIN-168 / JH642;

```

```

RX MEDLINE: 97175542.
RA Homuth G., Masuda S., Mogk A., Kobayashi Y., Schumann W.;
RT "The dnaK operon of Bacillus subtilis is heptacistronic.";
RL J. Bacteriol. 179:1153-1164(1997).
CC -1- FUNCTION: INTERACTS WITH DnaJ, TO DISASSEMBLE A PROTEIN COMPLEX
CC AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY
CC WITH GRE, THE APASE ACTIVITY OF DnaK (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DnaJ AND TO EUKARYOTIC
CC DnaJ-LIKE PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL: M64964; AAA2529.1; -
DB EMBL: D84432; BAA12465.1; -
DB EMBL: X52064; CAA36287.1; -
DB EMBL: D83717; BAA12077.1; -
DB EMBL: 259117; CAB14488.1; -
DB PIR: S09501; S09501.
DB PIR: S27506; S27506.
DB PIR: B41874; B41874.
DB HSSP: P08622; 1XBL.
DB SUBTILIST; BG10665; DnaJ.
DB PRINTS; PR00625; DnaJPROTEIN.
DB PROSITE; PS00636; DnaJ_1; 1.
DB PROSITE; PS0076; DnaJ_2; 1.
DB PROSITE; PS00637; DnaJ_CXXCXXGXG; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXXCXXGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
KW Chaperone; DNA replication; Heat shock; Repeat.
FT DOMAIN 3 71 DnaJ-LIKE.
FT DOMAIN 76 103 GLY-RICH.
FT REPEAT 145 152 CXXCXXGXG MOTIF.
FT REPEAT 162 169 CXXCXXGXG MOTIF.
FT REPEAT 188 195 CXXCXXGXG MOTIF.
FT REPEAT 202 209 CXXCXXGXG MOTIF.
SQ SEQUENCE 372 AA; 40847 MW; 353C0862F35A5A95 CRC64;

Query Match 17.1%; Score 293; DB 1; Length 372;
Best Local Similarity 33.3%; Pred. No. 1.4e-15;
Matches 90; Conservative 31; Mismatches 85; Indels 64; Gaps 10;

QY 3 DYEEVLGVORHASPEDIKAYRKLAKMHPDKNPENKEAEKPKQVAEAYEVLSDAKR 62
DB 5 DYEEVLGVSRSATADIRAKRLAKMHPDKNPENKEAEKPKQVAEAYEVLSDAKR 62
QY 63 DYDYKVG-----KEGLNGGGGGSHDFEFERFTFRNPDVFRFFFGGRDPFDFEDP 118
DB 63 AHYDQFGHTDPNQGRFGGGGFGGDFG---IGF-----DILFSSIFG----- 103
QY 119 FDFDFGNRRGRPRGSRSGRTGSAFSGFPFSGSGFSF-----DTGTFSGSLHGGL 172
DB 103 -----GTRRRDRLKARADLYQWTLSFEDAFAFKETIIEIPRENCETCGGAKFCT 157
QY 173 TSFSTSGSGMGNFKISTTKMNGR-----KITTKRIYENG-----QERV 216
DB 158 NPEITSHCGSGQLNVEQNTPRGKYVNRVCHNCBGTGKITTKKACDGGGKGIKKRKRI 217
QY 217 EYE-----EDGQLKSLT-----INGVAD 235
DB 218 NVTIPAGVDDGQQLRLTSGGEGINGGLPD 247

RESULT 14
DnaJ_CUCSA

```

	Query Match	16.7%	Score 287	DB 1:	Length 413;	
	Nearly Local Similarity	31.0%	Pred. 4.7e-15;			
	Matches	85;	Conservative	30;	Mismatches	73;
					Indels	86;
					Gaps	11;
OY	4 YVEVLGVORHASPEDIKAKAYRKRLALAKWHPDK--NPENKEEAEARKFEQVAAEALEYLSDAKK	61				
Db	13 YEELIGVSNNASQDCLKAKRYKAALKKNPKDGKDE-----KKRELGAQAEVLSDPDK	65				
OY	62 RDIITYGKEGLNG-GGGGSHFDSPFEFGFTFRNPDDVFRRFFGGGRDPFSDFEDDPPE	120				
Db	66 REIYYOYGEDALKEGMGGGGH-----DPDFIOFSGFG-SFGCG-----	106				
OY	121 DFGNRRGRG-----SRNRGTGSFFSAFSGF	147				
Db	106 ---GSSRGRRQRGEDVTHPLKVLSELDLYNGTSKLSTLSRNVIYCCKCKGSKGASAMKC	162				

RESULT	15			
ID	DNAJ_CLOAB	STANDARD:	PRF:	374 AA.
AC	P30725;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	DNAJ PROTEIN.			
GN	DNAJ.			
OS	Clostridium acetobutylicum.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
CC	Clostridium.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 94123950.			
RA	Behrens S., Narberhaus F., Bahl H.;			
RT	"Cloning, nucleotide sequence and structural analysis of the			
RT	Clostridium acetobutylicum dnaJ gene.";			
RL	FEBS Microbiol. Lett. 114:53-60(1993).			
RN	[2]			
RP	SEQUENCE OF 1-72 FROM N.A.			
RC	STRAIN-DSM 1731;			
RX	MEDLINE: 92250425.			
RA	Narberhaus F., Giebler K., Bahl H.;			
RT	"Molecular characterization of the dnaK gene region of Clostridium			
RT	acetobutylicum, including grpe, dnaJ, and a new heat shock gene.";			
RL	J. Bacteriol. 174:3290-3299(1992).			
CC	-1- FUNCTION: INTERACTS WITH DNAK, TO DISASSEMBLE A PROTEIN COMPLEX			
CC	AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY			
CC	WITH GRAPE, THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).			
CC	-1- INDUCTION: BY HEAT SHOCK.			
CC	-1- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC			
CC	DNAJ-LIKE PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X69050; CAA48792.1; -.			
DR	EMBL: M74569; AAA23247.1; -.			
DR	PIR: A41873; A41873.			
DR	HSSP: P25685; 1HDJ.			
DR	PRINTS: PR00625; DNAJPROTEIN.			
DR	PROSITE: PS00636; DNAJ_1; 1.			
DR	PROSITE: PS50076; DNAJ_2; 1.			
DR	PROSITE: PS00637; DNAJ_CXKXGXG; 1.			
DR	PFAM: PF00226; DnaJ; 1.			
DR	PFAM: PF00684; DnaJ_CXKXGXG; 1.			
DR	PFAM: PF01556; DnaJ_C; 1.			
KW	Chaperone; DNA replication; Heat shock; Repeat.			
FT	DOMAIN	3	72	DNAJ-LIKE.
FT	DOMAIN	76	108	GLY-RICH.
FT	REPEAT	150	157	CXKXGXG MOTIF.
FT	REPEAT	167	174	CXKXGXG MOTIF.
FT	REPEAT	193	200	CXKXGXG MOTIF.
FT	REPEAT	207	214	CXKXGXG MOTIF.
SO	SEQUENCE	374 AA;	40401 MW;	DDC6B58511F201E CRC64;

Query Match 16.78; Score 286.5; DB 1; Length 374;

Best Local Similarity 37.1%; Pred. No. 4.6e-15;
Matches 78; Conservative 30; Mismatches 67; Indels 35; Gaps 8;

```
OY 3 DYEVVGVRHASPEDIKAVRKALAKWHPDKNPENKEAEKFKQVAEATVLSDAKR 62
    |||||:::|:|||||:||||| | ||| |: |||::| |
Db 5 DYEVVLGLEGASDEIKAKRKLAKYHPDKNRGNK-EAEKFKINEAYOVLSDPDK 63
    ||::| |||| | || | | | | | |
OY 63 DYDRYKKGKGLNGCGGSHEDSPFEFGF-TERNPDVFREFFGGGRODPFSDFEFDPED 121
    ||::| |||| | || | | | | | |
Db 64 ANYDRFGTADFNCGGGEFD----FSGFGDRLGDLIFNSFFG--GFS----- 107
    ||::| |||| | || | | | | | |
OY 122 FFGNRGRGRSRKSGTGSFFSAFSGPPS--FG----SGFSSFDIGFTSFGSIGHGGLTSF 175
    ||::| |::| | | | | | | | | |
Db 107 -GGSSRRARKDAPQGRNDMEYSISLTFEBAVFGVEKSNITRSENCETCGTGAKKGTSPK 165
    ||::| |::| | | | | | | | | |
OY 176 SSTSRGSG-----MGNFKSISTSTK 196
    ||::| |::| | | | | | | | |
Db 166 TCDKCGGTGTIRVQRNTPLGSEVYIOSGCDK 195
    ||::| |::| | | | | | | | |
```

Search completed: June 30, 2000, 11:56:23
Job time: 7138 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 11:53:56 ; Search time 53.9 Seconds

(without alignments)
424.495 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716
Sequence: 1 MVDYEVGVGRHASPEDIK.....EAEKGVREEVDRQSLDRT 330Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1230	71.7	241	4 075190	075190 homo sapien
2	1158.5	67.5	242	11 054946	054946 mus musculu
3	1060	61.8	215	4 095806	095806 homo sapien
4	847.5	49.4	242	-11 035723	035723 mus musculu
5	363	21.2	336	5 076224	076224 trypanosoma
6	361.5	21.1	348	10 049457	049457 arabidopsis
7	358	20.9	338	5 076230	076230 trypanosoma
8	343	20.0	331	5 020774	020774 caenorhabdi
9	339	19.8	348	11 089114	089114 mus musculu
10	337	19.6	348	4 075953	075953 homo sapien
11	332.5	19.4	330	3 013303	013303 cryptococcu
12	332	19.3	178	4 013431	013431 homo sapien
13	330	19.2	392	2 084345	084345 chlamydia t
14	327	19.1	354	5 096123	096123 plasmodium
15	309	18.0	402	5 045502	045502 caenorhabdi
16	307.5	17.9	331	10 09X1F5	09X1F5 arabidopsi
17	303.5	17.7	392	2 0929E9	0929E9 chlamydia p
18	298	17.4	411	5 096455	096455 dictyosteli
19	295.5	17.2	420	10 09ZWK3	09ZWK3 salix gligl
20	291	17.0	369	2 025890	025890 helicobacte

21	288.5	16.8	328	5 096212	096212 plasmodium
22	287.5	16.8	383	2 087778	087778 lactobacill
23	285	16.6	369	2 09ZJ02	09ZJ02 helicobacte
24	283	16.5	415	10 09ZS26	09ZS26 hevea brasl
25	283	16.5	420	10 042530	042530 arabidopsis
26	283	16.5	420	10 022663	022663 arabidopsis
27	282	16.4	419	10 043177	043177 solanum tub
28	281	16.4	383	2 09XC46	09XC46 porphyromon
29	278.5	16.2	118	10 09ZRE0	09ZRE0 nicotiana t
30	276.5	16.1	381	3 09Y8G9	09Y8G9 schizosacch
31	276.5	16.1	386	3 094625	094625 schizosacch
32	275.5	16.1	371	2 09ZFC5	09ZFC5 methylovoru
33	274.5	16.0	419	10 065160	065160 zea mays (m
34	273.5	15.9	369	2 09WZV3	09WZV3 thermotoga
35	272.5	15.9	379	2 068797	068797 campylobact
36	268.5	15.6	423	10 024074	024074 medicago sa
37	265	15.4	379	2 052065	052065 pasteurella
38	263.5	15.4	385	2 087385	087385 vibrio harv
39	263.5	15.4	480	4 075472	075472 homo sapien
40	262.5	15.3	390	5 026953	026953 trypanosoma
41	261	15.2	316	2 09XC82	09XC82 rhodothermu
42	260	15.2	407	3 074752	074752 schizosacch
43	260	15.2	412	5 018427	018427 geodia cydo
44	260	15.2	540	5 077048	077048 plasmodium
45	259	15.1	372	2 085213	085213 campylobact

ALIGNMENTS

RESULT	ID	075190	PRELIMINARY;	PRT;	241 AA.
AC	075190;				
DT	01-NOV-1998 (Tremblrel. 08, Created)				
DT	01-NOV-1998 (Tremblrel. 08, Last sequence update)				
DT	01-NOV-1999 (Tremblrel. 12, Last annotation update)				
DE	MSJ-1.				
GN	MSJ-1 OR HSJ2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eukarya; Primates; Catarrhini; Hominoidea; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=TESTIS;				
RA	SAITO T., SEKI N.;				
RT	"A new member of human dnaJ-related gene family.";				
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	ZHANG W., WAN T., YUAN Z., CAO X.;				
RT	"HSJ2, a novel human homologue of the bacterial heat-shock protein DnaJ".				
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB014888; BAA32209.1; -.				
DR	EMBL; AF075601; AAD43194.1; -.				
DR	HSSP; P25685; 1HDJ.				
DR	PROSITE; PS00636; DnaJ_1; 1.				
DR	PFAM; PF00226; DnaJ; 1.				
KW	Heat shock.				
SO	SEQUENCE 241 AA; 26900 MW; D2F7F6BF CRC32;				
Query Match	71.7%; Score 1230; DB 4; Length 241;				
Best Local Similarity	100.0%; Pred. No. 1.1e-93;				
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	1 MVDYEVGVGRHASPEDIKAKYRLAKWHPDKNPKKEAEKRFKOVAAEYVLSDAK 60				
DB	1 MVDYEVGVGRHASPEDIKAKYRLAKWHPDKNPKKEAEKRFKOVAAEYVLSDAK 60				
OY	61 KRDIYDKRGKGLNGGGGGSHFSPPEFGFTFRNPDDVFRPFGGRDPFDFDPDFE 120				


```

Db 61 KRDIYDKYKKEGLNGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSFDEDPPE 120
QY 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 180
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 180
QY 181 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 231
    |||
Db 181 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 231

RESULT 2
054946 PRELIMINARY: PRT: 242 AA.
ID 054946
AC 054946;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MRJ.
GN MRJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129; TISSUE-TROPHOBLAST;
RA HUNTER P.J., SPANSON B.J., HAENDEL M., LYONS G.E., CROSS J.C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF035962; AAC16759.1; -.
DR HSSP: P25685; 1HDT.
DR PROSITE: PS00636; DNAJ_1; 1.
DR PFAM: PF00226; DnaJ; 1.
SQ SEQUENCE 242 AA; 27012 MW; AC627078 CRC32;

Query Match 67.5%; Score 1158.5; DB 11; Length 242;
Best Local Similarity 93.5%; Pred. No. 8.2e-88;
Matches 217; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPENKEEARKKQVAEAYEVLSDAK 60
    |||
Db 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPENKEEARKKQVAEAYEVLSDAK 60
QY 61 KRDIYDKYKKEGLNGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSFDEDPPE 119
    |||
Db 61 KRDIYDKYKKEGLNGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSFDEDPPE 120
QY 120 EDFGNNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 179
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 180
QY 180 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 231
    |||
Db 181 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 232

RESULT 3
095806 PRELIMINARY: PRT: 215 AA.
ID 095806
AC 095806;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DNAJ-LIKE 2 PROTEIN.
GN HSJ2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE: 99115663.
RA PEI L.;

```

```

RT "Plutitary tumor-transforming gene protein associates with ribosomal
RT protein S10 and a novel human homologue of DnaJ in testicular cells.";
RL J. Biol. Chem. 274:3151-3158(1999).
DR EMBL: AF080569; AAD16010.1; -.
DR HSSP: P25685; 1HDT.
DR PROSITE: PS00636; DNAJ_1; 1.
SQ SEQUENCE 215 AA; 24051 MW; 49385BFB CRC32;

Query Match 61.8%; Score 1060; DB 4; Length 215;
Best Local Similarity 97.5%; Pred. No. 8.7e-80;
Matches 197; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPENKEEARKKQVAEAYEVLSDAK 60
    |||
Db 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPENKEEARKKQVAEAYEVLSDAK 60
QY 61 KRDIYDKYKKEGLNGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSFDEDPPE 120
    |||
Db 61 KRDIYDKYKKEGLNGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSFDEDPPE 120
QY 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 180
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 180
QY 181 GSGGMGNFKSISTSTKMVNGRK 202
    |||
Db 181 GSGGMGNFKSISTSTKMVNGRK 202

RESULT 4
035723 PRELIMINARY: PRT: 242 AA.
ID 035723
AC 035723;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TESTIS SPECIFIC DNAJ-HOMOLOG.
GN HSJ3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 98189155.
RA BERRUTI G., PEREGO L., BORGONOVO B., MARENGANI E.;
RT "MSJ-1, a new member of the DnaJ family of proteins, is a male germ
RT cell-specific gene product.";
RL Exp. Cell Res. 239:430-441(1998).
DR EMBL: U95607; AAC13944.1; -.
DR HSSP: P25685; 1HDT.
DR MGD: MGI:1306822; HsJ3.
DR PROSITE: PS00636; DNAJ_1; 1.
DR PFAM: PF00226; DnaJ; 1.
SQ SEQUENCE 242 AA; 26678 MW; D3DCA2B7 CRC32;

Query Match 49.4%; Score 847.5; DB 11; Length 242;
Best Local Similarity 70.0%; Pred. No. 2.8e-62;
Matches 163; Conservative 30; Mismatches 37; Indels 3; Gaps 3;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPENKEEARKKQVAEAYEVLSDAK 60
    |||
Db 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDKNPENKEEARKKQVAEAYEVLSDAK 60
QY 61 KRDIYDKYKKEGLNGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSFDEDPPE 118
    |||
Db 61 KRDIYDKYKKEGLNGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSFDEDPPE 120
QY 119 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 178
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGLTSSSTSF 179

```


DR PFAM; PF01556; DnaJ_C; 1.
KM Heat shock.
SQ SEQUENCE 348 AA; 39133 MW; 92AF8A03 CRC32;

Query Match 19.6%; Score 337; DB 4; Length 348;
Best Local Similarity 32.2%; Pred. No. 3.8e-20;
Matches 96; Conservative 39; Mismatches 101; Indels 62; Gaps 11;

QY 3 DYEVVLGVQRHASPEDIKKAYRKIALKWPDKNPENKEEAKRKQVAEAYEVLSDAKR 62
D 4 DYVKILIPSGANDEIKKAYRKMAKLYHPDKNKE--PNAEEKKEIAEAYDVLSDPKR 61
QY 63 DYVKKYKEGNGGGG--GSHFDSPEFEGFTR-NPDVDYREFEGGDDPSFDFEDPE 119
D 62 GLVDYQEEGLTKTGSGG--SSGSFHYTFHGDPAATPASFPGG-----SNPF 108
QY 120 EDFGNRGRPGRSRSGTGSFSAFSGFSFGSFFDTGTFSGLSHGILGTFSSSTS 179
D 109 DIFPASSRSTR-----PFGSFDDPDDMDVDEDEDFGAFGRGFGNLSRGPRRA 156
QY 180 FGGSGMGNFNSISTSTKMWNGRKITTKRIYENGQERVEVEE--DGQLKSLTI----- 230
D 157 -----PEELPRRKVQDDPPVYH--ELRVSLSEIYHGSTRKMKITRRRLNPD 200
QY 230 -NGVADDDALXEEHRMG-----QNVLPAPAGLRPPRPAPSLIRKXPHCLSKKEG 281
D 201 GRVTRTEDKILHIVIKRGWKEGTRITEPKEGDAPDPNIPADIVELDKRPAHRRRDG 258

RESULT 11

013303 PRELIMINARY; PRT; 330 AA.

AC 013303;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE S1S1 PROTEIN.
GN S1S1.
OS Cryptococcus curvatus.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellales;
OC anamorphic Tremellales; Cryptococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 20509;
RX MEDLINE; 98220306.
RA SPECHT V., LUBECK M., KINDL H.;
RT "Heat shock transiently enhances the synthesis rate of Sisd, a
RT ribosome-associated DnaJ protein in the oleagenous yeast *Apiotrichum*
RT curvatum";
RL Yeast 14:419-430(1998).
DR EMBL; Y12080; CAA72798.1; -.
DR HSSP; P25685; 1HDJ.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PFAM; PF00226; DnaJ; 1.
SQ SEQUENCE 330 AA; 34596 MW; 4FD216DA CRC32;

Query Match 19.4%; Score 332.5; DB 3; Length 330;
Best Local Similarity 32.8%; Pred. No. 8.3e-20;
Matches 97; Conservative 33; Mismatches 78; Indels 89; Gaps 13;

QY 3 DYEVVLGVQRHASPEDIKKAYRKIALKWPDKNPENKEEAKRKQVAEAYEVLSDAKR 61
D 6 EYVKTLLSKRSPADIKKAYRKESLKWHPDKNGDKRATAEKKFKKGAAYEVLSDPEK 65
QY 62 RDIDYKXGKBELN-----GGGGGSHDSDPFEFG-----PFRNPDVDFEFEGGR 107
D 66 RKIIDQGEERLKGMPAGGGGGGPGFSS--FGAGGAAPASPTPTDNDIFNAFF-- 120
QY 108 DPFSDFFEDPFDFFGNRRGRPGRSRSGTGSFSAF--SGPFSFGSGSFFDTGTFSG 165
D 120 -----SQMGSGGGGGGDEYAFSGCGGPGGGRTRRRRGRGGMGMG 159

QY 166 SL-GHGILTSSTSFSGSG-----MGNFKSISTS 194
D 160 GMGGMGMPGMG--GMGGMGMPGGRYARPEPTPAETIERPLPLSELYKGGTRLRIT 217
QY 195 TKMVGRKITTTRKIYENGQERVEVEEDQL-KSLTINGVADDDALXEEHRMGONV 249
D 218 RHRNRGR--TNEKILE-----VAIKPGWAKGTKYTFAGAGNED-----EYGRQNI 261

RESULT 12

013431 PRELIMINARY; PRT; 178 AA.

AC 013431;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HEAT SHOCK PROTEIN HSP40 HOMOLOG.
GN DNaJw.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER.
RA WON M., MOON K.M., LEE C.E., YOO H.S.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; U40992; AAC14483.1; -.
DR EMBL; U41290; AAB07346.1; -.
DR HSSP; P25685; 1HDJ.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PFAM; PF00226; DnaJ; 1.
KW Heat shock.
SQ SEQUENCE 178 AA; 19930 MW; 5A1337E4 CRC32;

Query Match 19.3%; Score 332; DB 4; Length 178;
Best Local Similarity 49.4%; Pred. No. 4.2e-20;
Matches 76; Conservative 16; Mismatches 40; Indels 22; Gaps 6;

QY 3 DYEVVLGVQRHASPEDIKKAYRKIALKWPDKNPENKEEAKRKQVAEAYEVLSDAKR 62
D 4 DYICLTIGEKASDEDIKKAYRKQALKYHPDKN--KSQLAEKKEKREVAEAYEVLSDPKR 61
QY 63 DYDKYKREGNGGGGGSHDSDPFEFGFTRNPDVDFEFEGGGRDPSFDFEDPFD 122
D 62 EIVDQFGEGLKKGAGGTDGCGGFRYTF-HGDPHATPAAFKXGSPFXI-----F 111
QY 123 FGNRRGPRGSRSG-----TGSFSAF-----SGFP 148
D 112 FGRRWG--GGRDSEMEIDGDPFSAFGFSMNGYP 143

RESULT 13

084345 PRELIMINARY; PRT; 392 AA.

AC 084345;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HEAT SHOCK PROTEIN J.
GN DNaJ.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., PAN J., MARATHE R., ARAYIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis";
RL Science 0:0-0(1998).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/UM-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.P., LINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE001307; AAC67936.1; -.
 DR HSSP; P08622; 1XBL.
 DR PROSITE; PS00636; DNAJ_1; 1.
 DR PFAM; PF00226; DnaJ; 1.
 DR PFAM; PF01556; DnaJ_C; 1.
 DR PFAM; PF00684; DnaJ_CXXCXXG; 1.
 DR PRINTS; PR00625; DNAJPROTEIN.
 KW Heat shock.
 SQ SEQUENCE 392 AA; 41916 MW; 8D33F656 CRC32;

```

QY      2 VDYVEVGVORHASPEDIKKAYKRLKALKNPDKNPENKKEAEKRFQVAEAEVYLSDAKK   61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 MDYITLIGVAKTAPPEIKKAYKRLKAVKHYIPDKNP--GDMAERERFEVSEAEVIGDAQK   59

QY      62 RDIDYKYGKBEGLNGGGGGSHFDSPEFEGFTERNPDVFRFEFGADPSPFD-----F   114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      60 RESYDRYCKGKGFPGAGCGFG-----GAGMGMMDALRTFMGA---FGGGFGGGGFG   108

QY      115 FEDPFF-----EDPEFGNRPRPGRSRSG-----TGSPFSAFSGPFS--FGSGFSSFDGTFS   163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      109 FEGFGGLGEAFGMRGSESSRQAGSKKHVITLTFEAAKGYVEKELLYSGYKSCD-----   164

QY      164 FGSJLHGGLTFSSTSG-----GSGN-----GNFKSISTYKMV--NGRITTKRIIV   209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      164 -ACGSGS-----ANTAGVYVCDRCCKSGGYVSRFFGSMASCTPCDSCGEGVITDPCSY   217

QY      210 ENGGERVEVEEDGLKSLTINGVAADDALCKEEDMR-----RGOWVLPA   252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      218 CRGGGRIT-----DKRSVHNIRPAGVDSGRMLKMEGYGADGANGAP   259

```

RESULT	14	
096123		
ID	096123	PRELIMINARY; PRT; 354 AA.
AC	096123;	
DT	01-MAY-1999 (TrEMBLrel. 10, Created)	
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)	
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)	
DE	PROTEIN WITH DNAJ DOMAIN.	
GN	PF0090C.	
OS	Plasmodium falciparum.	
CC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 99021743.	
RA	GAANDER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,	
RA	KOONIN E.V., SHILLON S., MASON T., YU K., FUJII C., PEDERSON J.,	
RA	SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,	
RA	SAULZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,	
RA	FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;	
RT	"Chromosome 2 sequence of the human malaria parasite Plasmodium	
RT	falciparum.";	
RL	Sequence 282:1126-1132(1998) .	
DR	EMBL; AE001370; AAC71808.1; -	
DR	HSSP; P08622; 1XBL.	
DR	PROSITE, PS00636; DNAJ_1. 1.	
QD	SEQUENCE 354 AA; 40329 MW; 35500567 CRC32;	

Query Match	19.1%;	Score 327;	DB 5;	Length 354;
Best Local Similarity	37.6%;	Pred. No. 2.6e-19;		

[illegible]

RESULT	15		
045502		PRELIMINARY:	PRT: 402 AA.
ID	045502		
AC	045502:		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)		
DE	F39B2.10 PROTEIN.		
CN	F39B2.10.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilita;		
OC	Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	DOBSON R.;		
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 94150718.		
RA	WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,		
RA	BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,		
RA	GRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,		
RA	GARNER A., GREEN P., HAMKINS T., HILLIER L., JIER M., JOHNSTON L.,		
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER L., LATREILLE P.,		
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,		
RA	PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,		
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUTTON J.,		
RA	THERRER-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,		
RA	WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOLDMAN P., III OF C.		
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RL	Nature 368:32-38(1994).		
CC	-1- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC DNAJ-LIKE		
CC	PROTEINS.		
DR	EMBL: 292834; CAB07390.1; -.		
DR	HSPB, P25685; 1HDJ.		
DR	PROSITE: PS00636; DNAJ_1; 1.		
DR	PROSITE: PS00637; DNAJ_CXXCXGXG; 1.		
DR	PFAM: PF00226; DnaJ; 1.		
DR	PFAM: PF00684; DnaJ_CXXCXGXG; 1.		
DR	PFAM: PF01556; DnaJ; 1.		
DR	PRINTS: PR00625; DNAJPROTEIN.		
KW	Chaperone; DNA replication.		
SO	SEQUENCE 402 AA; 44308 MW; 592FBB9CF CRC32;		

Query Match	18.0%	Score 309	DB 5	Length 402
Best Local Similarity	29.4%	Pred. No. 9	1e-18	
Matches	91	Conservative	38	Mismatches 64, Indels 116, Gaps 11
QY	4	YEVVLGVQVHNSPEDIKKIKRLALAKMHPDDKPNKEEAKRFQVLAENVLSDAKKRD	63	
DB	7	YVDVGVKFRDASDNLKKYKRRKAKLPHDDKPDGEO---FKQISQAEVYLSDDKKQ	62	

THIS PAGE BLANK (USPTO)

SEQUENCE 419 AA; 46700 MW; 66DB838A CRC32;

Query Match 28.4%; Score 534; DB 10; Length 419;
Best Local Similarity 36.7%; Pred. No. 6.5e-36;
Matches 130; Conservative 66; Mismatches 112; Indels 44; Gaps 13;

QY 26 FYKILGVPKRSIKDKKAYRKALQLHPDRNPDDPOAEKFFODLGAAYEVLSDSEKRRQ 85
D 14 YFEILGVKTAQEDLKKAYRKALIKNHPDKGD---PEKFKELAQAYEVLSDPEKREI 69
QY 86 YDTGEEGLKRGHSSHC-----DIFSHFPG--DFGFMFGGTPROQDRNPGRSDIIVDL 138
D 70 YDQIGEDALKEGSGGGGSHDPDIFSSFFGSPFG---GGGSSRRGRGRGEDVYHPL 126
QY 139 EYTLVEEYAGNF--VEYVRN-----KPVARQAPGRKCN--CR--QEMRTT--OLGPGRFQ 186
D 127 KVSLEDLYNGSKSLSLSRNVLSCKKSGKSGASKSGSCGCGSGMKVYTRQDGPSPMIQ 186
QY 187 MTQ-----EVV-----CDECPNVKLVNERTELEVEIPGVDRDMEYPIGEGEPH 231
D 187 QMHPHCNECKGTGEMINDKRCGCGCKEYVQEKVLEVYVEKGMONGQKITTFEGADEA 246
QY 232 VDGPGLRFRIRIKYVKHPIFERGGDLYTNVTVSLVESLVGFEMDITHLGKHYHT--SR 289
D 247 PDVTVGDIVFLQCKHDPKFRKGGDLFEVHTLSLTALCGFQVLIHLDRQLLISQPG 306
QY 290 DKITRPGAXXMKKEGLPNFDNNNIKSLITFEDVDPKQLEAREGIRKOLL 343
D 307 GEVVKPDQFKAINDGEMPMYORPFRMRGKLYIHFTVERP--DLSLSPDMCKALEAVLPRAVSQ 359

RESULT 7
Q92WK3 PRELIMINARY; PRT; 420 AA.
AC Q92WK3
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE DNJ HOMOLOG.
GN SGJ3.
OS Salix gligiana.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eucolecyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Salix.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FLOWER;
RA FUTAMURA N., ISHII-MINAMI N., HAYASHIDA N., SHIMOHARA K.,
RT "Expression of genes for proteins homologous to the bacterial
RT molecular chaperone DnaJ and DnaK in the Japanese willow (Salix
RT gligiana Seemen)."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAs, AND TO EUKARYOTIC DNAs-LIKE
CC PROTEINS.
DR EMBL: AB015601; BAA35121.1; -.
DR HSSP: P25685; 1HDJ.
DR MENDEL: 35038; Salgi.1253;35038.
DR PROSITE: PS00636; DNJ_1; 1.
DR PROSITE: PS00637; DNJ_CXXCXGXG; 1.
KW Chaperone; DNA replication.
SQ SEQUENCE 420 AA; 46563 MW; 84A45244 CRC32;

Query Match 27.9%; Score 526; DB 10; Length 420;
Best Local Similarity 35.8%; Pred. No. 2.9e-35;
Matches 129; Conservative 65; Mismatches 124; Indels 42; Gaps 12;

QY 26 FYKILGVPKRSIKDKKAYRKALQLHPDRNPDDPOAEKFFODLGAAYEVLSDSEKRRQ 85
D 14 YFEILGVKTAQEDLKKAYRKALIKNHPDKGD---PEKFKELAQAYEVLSDPEKREI 69
QY 86 YDTGEEGLKRGHSSHC-----SHG--DIFSHFPGDFGFMFGGTPROQDRNPGRSDIIVDL 139

D 70 YDQIGEDALKEGSGGGSHDPDIFQSFPG--GNPFGGSGSSRRGRGRGEDVYHPLK 128
QY 140 VTLVEYTAGNF--VEYVRN-----KPVARQAPGRKCN--NCR--QEMRTTOLGPGRFQ 187
D 129 VSEFDLYNGTSKSLSLSRNVLSCKKSGKSGASKSGSCGCGSGMKVSIHRLGSPMIQ 188
QY 188 TOEVV-----CDECPNVKLVNERTELEVEIPGVDRDMEYPIGEGEPH 232
D 189 QMHPHCNECKGTGETINDKRCPOCKGKGVQEKVLEVYVEKGMONGQKITTFEGADEAP 248
QY 233 DGPGLRFRIRIKYVKHPIFERGGDLYTNVTVSLVESLVGFEMDITHLGKHYHSD-- 291
D 249 PDVTVGDIVFLQCKHDPKFRKGGDLFEVHTLSLTALCGFQVLIHLDRQLLISQPG 308
QY 291 KITRPGAXXMKKEGLPNFDNNNIKSLITFEDVDPKQLEAREGIRKOLL--KQSVQ 349
D 309 EVVKPDQFKAINDGEMPMYORPFRMRGKLYIHFTVERP--DLSLSPDMCKALEAVLPRAVSQ 367

RESULT 8
O65160 PRELIMINARY; PRT; 419 AA.
ID O65160
AC O65160;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE DNJ-RELATED PROTEIN ZMDJ1.
GN MDJ1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RA BASZCZYNSKI C.L., BARBOUR E., ZEKA B., MADDOCK S.E., SWENSON J.L.;
RL Maydis 42:189-201(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA BASZCZYNSKI C.L.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAs, AND TO EUKARYOTIC DNAs-LIKE
CC PROTEINS.
DR EMBL: AF053468; AAC08009.1; -.
DR HSSP: P25685; 1HDJ.
DR MENDEL: 29613; Zeama.1253;29613.
DR PROSITE: PS00636; DNJ_CXXCXGXG; 1.
DR PROSITE: PS00637; DNJ_1; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXXCXGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
DR PRINTS: PR00625; DNJPROTEIN.
KW Chaperone; DNA replication.
SQ SEQUENCE 419 AA; 46725 MW; BE97EBB2 CRC32;

Query Match 27.3%; Score 515; DB 10; Length 419;
Best Local Similarity 35.8%; Pred. No. 2.3e-34;
Matches 125; Conservative 61; Mismatches 115; Indels 48; Gaps 10;

QY 26 FYKILGVPKRSIKDKKAYRKALQLHPDRNPDDPOAEKFFODLGAAYEVLSDSEKRRQ 85
D 14 YFEILGVKTAQEDLKKAYRKALIKNHPDKGD---PEKFKELAQAYEVLSDPEKREI 69
QY 86 YDTGEEGLKRGHSSHC-----GDIESHFPG--DFGFMFGGTPROQDRNPGRSDIIVDL 139
D 70 YDQIGEDALKEGSGGGSHDPDIFSSFFGSPFG---GGGSSRRGRGRGEDVYHPLK 126
QY 140 VTLVEYTAGNF--VEYVRN-----KPVARQAPGRKCN--NCR--QEMRTTOLGPGRFQ 185
D 127 VSLLEDLYNGTSKSLSLSRNVLSCKKSGKSGASKSGSMRCGCGSGMKVYTR--QLGFSMT 184
QY 186 QMTGEVVCDECPNV-----KLVNERTELEVEIPGVDRDMEYPIGEGE 229

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, RANGELANDER;
 RA FRIGIS G., MELE G., GIANNINO D., MARIOTTI D.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA FRIGIS G., MELE G., GIANNINO D., MARIOTTI D.;
 RT "Isolation and characterization of a DnaJ-like gene from *alfalfa*."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC DNAJ-LIKE
 CC PROTEINS.
 DR EMBL: AF000995; CA04447.1; -.
 DR EMBL: AF069507; AAC19391.1; -.
 DR HSSP: P25685; 1HDJ.
 DR MENDEL: 26909; Medea; 1253; 26909.
 DR PROSITE: PS00637; DNAJ_CXXCXGXG; 1.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PFAM: PF00226; DnaJ; 1.
 DR PFAM: PF00684; DnaJ_CXXCXGXG; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 DR PRINTS: PR00625; DNAJPROTEIN.
 KW Chaperone; DNA replication.
 KM SEQUENCE 423 AA; 47197 MW; 7BB3C880 CRC32;

Query Match 27.0%; Score 507.5; DB 10; Length 423;
 Best Local Similarity 35.8%; Pred. No. 9, 6e-34;
 Matches 121; Conservative 57; Mismatches 117; Indels 43; Gaps 9;

QY 26 FYKLGPVRSASIDIKAYRKALALQHPDRNPDPQAOEKFODLGAAYVLSDESKRQ 85
 DB 15 YVDLGVSKASASEIKKAYRKAKMKNHPDKGDP---PEKFKELGAYVLSPEKKEL 70
 QY 86 YDVFGEGLKDGQSHG-----DIFSHFGDFMGFTGRQDORNIIPRSDIIVDL 138
 DB 71 YDQYGEALKEHMGOGGSGSFHNPDIFFESFG-AGFGGGGPRSA--RROKQSGDVVHSI 127
 QY 139 EYVLEEVYAGNFVEY-VRNKPVARAQAPGK-----RKCNCR---OEMRTTOLGPRFQ 186
 DB 128 KVSLEEDVYNGTTRKLSLRNALCSKCKSGSGTAGRCFGCGTGKIRRDIGLGMIO 167
 QY 187 MPEYV-----CDECPNVKLVNEERTLEVTEPGVDRDMETPRIGEGEPH 231
 DB 188 OMQHVCPDCKGTGEVISERDCRQCKGNKKTQEKVLEVHEVKGMOGKHIVPEGADDEL 247
 QY 232 VDPEPGDLRPRIRIKVPHPIERGGDLYTNVTVSLVESLGFEMDITHLGHVHLSRD- 291
 DB 248 PDITIGDIYFVLYQKHPRKRRERDHLIHNLSTLDALCGFQFNVTILDGROLVLSNP 307
 QY 291 -KITRPGAXXMKKGEGLPNFDNNNKGSLIITFDVDFP 327
 DB 308 GEVIRKPGOHKAINDEGMPOHGRPFMKGRILYKFSVDFP 345

RESULT 12
 ID 016303 PRELIMINARY; PRT; 439 AA.
 AC 016303;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE T05C3.5 PROTEIN.
 GN T05C3.5.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditidae; Rhabditidae; Peleoderinae; *Caenorhabditis*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BRKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTWING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SVALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 RT *elegans*."
 RL Mature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA BLANCHARD M., BRADSHAW H., KRAMER J.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC DNAJ-LIKE
 CC PROTEINS.
 DR EMBL: AF016428; AAB65361.1; -.
 DR HSSP: P25685; 1HDJ.
 DR PROSITE: PS00637; DNAJ_CXXCXGXG; 1.
 DR PFAM: PF00226; DnaJ; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 DR PFAM: PF00684; DnaJ_CXXCXGXG; 1.
 DR PRINTS: PR00625; DNAJPROTEIN.
 KW Chaperone; DNA replication.
 KM SEQUENCE 439 AA; 47514 MW; B3453315 CRC32;

Query Match 26.3%; Score 495; DB 5; Length 439;
 Best Local Similarity 35.0%; Pred. No. 1, 1e-32;
 Matches 129; Conservative 56; Mismatches 116; Indels 68; Gaps 14;

QY 27 YKILGVPVRSASIDIKAYRKALALQHPDRNPDPQAOEKFODLGAAYVLSDESKRQ 86
 DB 15 YTLNVRPDAQSADIKSYFKLAKETHPNKPDH---GDKFEISFAVEVLSPEKKRLY 71
 QY 87 DTYGEGELKDGQSHG-----HGDFSHFG-----DFGFMFGT----- 121
 DB 72 DARGLEGVGGGAGGGGGGPGGLFHFEGGAGGDDDDDDMGHPRFGGLGNGKMR 131
 QY 121 --PRODRNIPRGSIDIVLEVTLEVYAGNFVEY-VRNKPVARAQAPG-----KRCN 170
 DB 132 GGFRRR---KFQDTVHPLNVTLELVYGKTSKLSKALCKTCGSGGKRGKRYCD 186
 QY 171 -CR-OEMRT--TOLGPRFOMTOEVVCECP-----NKLTVNEERTLEV 210
 DB 187 ACRGRGVKTIYQDILGFMGLQMO-VHCDACKSGSGKVPAGDKCKCHGEKEVSKITLEV 245
 QY 211 EIEPGVDRDMETPRIGEGEPHVDPEPGDLRPRIRIKVPHPIERGGDLYTNVTVSLVES 269
 DB 246 HVLPGKHNDKITFGKGDQSDPDGPDGVVTVIQKHODIKRGGDDLHMFKKSLNWA 305
 QY 270 LVGFEMDITHLGHVHLS-RDKITRPGAXXMKKGEGLPNFDNNNKGSLIITFDVDFP 327
 DB 306 LCGYNELTKHLDGHLVLSKGDVAKPDVITGVLGKGMPPNKKYDELKGNLFEVEVEFP 365
 QY 328 KEQLTEAR 336
 DB 366 KEHFLDDEK 374

RESULT 13
 ID 066921 PRELIMINARY; PRT; 376 AA.
 AC 066921;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
 DE CHAPERONE DNAJ.
 GN DNAJ.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VE5;
 RX MEDLINE: 98196666.
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus";
 RL Nature 392:353-358(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VE5;
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A6000703; AAC06881.1; -
 DR HSSP: P08622; 1XBL.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PFAM: PF00226; DNAJ_1; 1.
 DR PFAM: PF01556; DNAJ_C_1.
 DR PFAM: PF00684; DNAJ_CXXGXG; 1.
 DR PRINTS: PR00625; DNAJPROTEIN.
 SQ SEQUENCE 376 AA; 41963 MW; 07BC58EB CRC32;

Query Match 26.0%; Score 490; DB 2; Length 376;
 Best Local Similarity 34.5%; Pred. No. 2.2e-32;
 Matches 128; Conservative 61; Mismatches 118; Indels 64; Gaps 15;

QY 24 RDYKILGVRSASIKDIKAYRKRLALQHPDNPPDPOAQEKFODLGAGAYEVLSDSEK 83
 DB 7 KDYELIGVPRMSOEIKKAYRKYKHPD-ICKKPECEKEKELNEAYOVLSDSEK 65
 QY 84 KQYDTYGEELKDGHOSSHG-----DIFSHFPG-DEGFME---GGTPRQ 123
 DB 66 KLVDMV-----GHAFAEGAGAAQARVETLEIRPIEILNEFPDFDGLSIFERATGRRRA 118
 QY 124 QDNPIRGSDIIVDELYTEVTAGNFVEYVRNKPVARQAPG-----KRC-NCROE 174
 DB 119 RRRRSYGEIIVVPEITTEAEKGTVPLEVEREYPCSAAGGTGYDESKSRICPTCGGR 178
 QY 175 MRTTQGGPRFQMTQEVVDCPCNVKLVNE-----ERTLEVELEPGVRGME 221
 DB 179 GEFVQ-GNMFQVQR--TCPTCGEGEVIYENCHACTGRTGYGLVKEITKVIIPGVRGSK 235
 QY 222 YPTIGEGEPV---DEPGDLRFRIKVKVHPHIFERRGDDLYTNVTVSLVSLVGFEMDIT 278
 DB 236 --LVNEGKAGAGYGGPRGLYIIVKPKRIEPRKGDLDYVNIITYPPAVLSTGEVPR 293
 QY 279 HLDGKHVHSRDKITRPGAXXMKKGEGLPNFDNNNIKSLITFDVDPK---EQLTEE 334
 DB 294 TLDEGKKVVIIPGTEKELIKVPGKMPRLKSGG-RGDLVVRVHIDVPRIGVLSKILGD 352
 QY 335 ARGIGKOLLKO 345
 DB 353 GKK-VRELLKO 362

RESULT 14
 ID 020774 PRELIMINARY; PRT; 331 AA.
 AC 020774;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
 DE F54D5.8 PROTEIN.

GN F54D5.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA COLES L., MATTHEWS L.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU 2., DURBIN R., FAYELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAKINS T., HILTER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LAVREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFEEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans";
 RL Nature 368:32-38(1994).
 DR EMBL: Z66513; CA91334.1; -
 DR HSSP: P25685; 1HDV.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PFAM: PF00226; DNAJ_1; 1.
 DR PFAM: PF01556; DNAJ_C_1.
 SQ SEQUENCE 331 AA; 36265 MW; E00F46D5 CRC32;

Query Match 25.0%; Score 470.5; DB 5; Length 331;
 Best Local Similarity 32.4%; Pred. No. 7.3e-31;
 Matches 119; Conservative 55; Mismatches 102; Indels 91; Gaps 9;

QY 23 GRDYKILGVRSASIKDIKAYRKRLALQHPDNPPDPOAQEKFODLGAGAYEVLSDSEK 82
 DB 2 GKDYKVLGSKGATWDEIKKAYRKALKYHPDN-KEAGLENKFKEIAEYDVLSDSK 60
 QY 83 KQYDTYGEELKDG-----HOSSHG---IFSHFPG----- 113
 DB 61 KKITDQGESEBLKGGGAGAGGGGGMHYERGDPMNIFSSFGSDPRFAGCGMEDLG 120
 QY 113 -----FGMEGTPROQDNIRGSDIIVDELYTEVTAGNFVE 152
 DB 121 GGAGGPMFMFNGGMDMGFMGNGGRRGHARQDPA---VLADLSVSLSEDLVGK--- 174
 QY 153 VVRNKPVARQAPRKRCNCRQEMTQDLGPRFQMTQEVVDCPCNVKLVEERTLEVEI 212
 DB 174 -----TTR---KMKITRKVMDNQR---EDKVLVTI 201
 QY 213 EPGVROGMEYFIEGEPHYDEPGDLRFRIKVKVHPHIFERRGDDLYTNVTVSLVSLVGF 272
 DB 202 KPGKSGTKITFRPESGQHPRTADIVFYKDKPHKFKREGSDIRKVEKISLSKALTG 261
 QY 273 FEMDITLHDKHVISRDKITRPGAXXMKKGEGLPNFDNNNIKSLITFDVDPKDLT 332
 DB 262 LDIMIPLDGADYRLQNDVIKPGTTRRLTGKGLPNKPSPSNRDILLIEFVEPR-SQLN 320
 QY 333 EAREGI 339
 DB 321 PTOREVI 327

RESULT 15
 ID 092FC5 PRELIMINARY; PRT; 371 AA.
 AC 092FC5;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
 DE PUTATIVE DNAJ.

